(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 August 2003 (14.08.2003)

PCT

(10) International Publication Number
WO 03/066086 A2

(51) International Patent Classification[7]: A61K 38/45, 31/70, A61P 3/00

(21) International Application Number: PCT/EP03/01094

(22) International Filing Date: 4 February 2003 (04.02.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

02002548.2	4 February 2002 (04.02.2002)	EP
02002707.4	6 February 2002 (06.02.2002)	EP
02002891.6	8 February 2002 (08.02.2002)	EP
02003748.7	19 February 2002 (19.02.2002)	EP
02004667.8	28 February 2002 (28.02.2002)	EP
02022101.6	2 October 2002 (02.10.2002)	EP

(71) Applicant *(for all designated States except US)*: **DEVELOGEN AKTIENGESELLSCHAFT FÜR ENTWICKLUNGSBIOLOGISCHE FORSCHUNG** [DE/DE]; Rudolf-Wissell-Strasse 28, 37079 Göttingen (DE).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **EULENBERG, Karsten** [DE/DE]; Am Graben 10b, 37079 Göttingen (DE). **BRÖNNER, Günter** [DE/DE]; Springstrasse 54, 37077 Göttingen (DE). **STEUERNAGEL, Arndt** [DE/DE]; Am Kirschberge 4, 37085 Göttingen (DE). **MEISE, Martin** [DE/DE]; An der Tränke 10, 37079 Göttingen (DE). **HÄDER, Thomas** [DE/DE]; Wiesenstrasse 17, 37073 Göttingen (DE).

(74) Agents: **WEICKMANN & WEICKMANN** et al.; Postfach 860 820, 81635 München (DE).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:
— *of inventorship (Rule 4.17(iv)) for US only*

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/066086 A2

(54) Title: PROTEINS INVOLVED IN THE REGULATION OF ENERGY HOMEOSTATIS

(57) Abstract: The present invention discloses casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, and/or homologous proteins regulating the energy homeostasis and the metabolism of triglycerides, and polynucleotides, which identify and encode the proteins disclosed in this invention. The invention also relates to the use of these sequences in the diagnosis, study, prevention, and treatment of metabolic diseases and disorders.

Proteins involved in the regulation of energy homeostasis

Description

This invention relates to the use of nucleic acid sequences encoding casein kinase 1 gamma (CSNK1G), GABA(A) receptor-associated protein (GABARAP), proliferation-associated 2G4 protein, 38kDa (PA2G4, also referred to as methionyl aminopeptidase homologous protein), molybdenum cofactor synthesis-step 1 protein (MOCS1), cell division cycle 10 protein homolog (CDC10, also referred to as septin and septin 7), pyruvate kinase (PK), calreticulin (CALR), or homologous proteins, and the polypeptides encoded thereby and to the use of these sequences or effectors thereof in the diagnosis, study, prevention, and treatment of diseases and disorders related to body-weight regulation, for example, but not limited to, metabolic diseases such as obesity as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, and gallstones.

Obesity is one of the most prevalent metabolic disorders in the world. It is a still poorly understood human disease that becomes more and more relevant for western society. Obesity is defined as an excess of body fat, frequently resulting in a significant impairment of health. Cardiovascular risk factors like hypertension, high blood levels of triglycerides and fasting glucose as well as low blood levels of HDL cholesterol are often linked to obesity. This typical cluster of symptoms is commonly defined as "metabolic syndrome" (Reaven, 2002, Circulation 106(3): 286-8 reviewed). The metabolic syndrome often precedes the development of type II diabetes and cardiovascular disease (McCook, 2002, JAMA 288:2709-2716). Besides severe risks of illness such as diabetes, hypertension and heart disease, individuals suffering from obesity are often isolated socially. Human obesity is strongly influenced by environmental

and genetic factors, whereby the environmental influence is often a hurdle for the identification of (human) obesity genes. Obesity is influenced by genetic, metabolic, biochemical, psychological, and behavioral factors. As such, it is a complex disorder that must be addressed on several fronts to achieve lasting positive clinical outcome. Obese individuals are prone to ailments including: diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, and gallstones.

Obesity is not to be considered as a single disorder but a heterogeneous group of conditions with (potential) multiple causes. Obesity is also characterized by elevated fasting plasma insulin and an exaggerated insulin response to oral glucose intake (Koltermann, J. Clin. Invest 65, 1980, 1272-1284) and a clear involvement of obesity in type 2 diabetes mellitus can be confirmed (Kopelman, Nature 404, 2000, 635-643).

Even if several candidate genes have been described which are supposed to influence the homeostatic system(s) that regulate body mass/weight, like leptin, VCPI, VCPL, or the peroxisome proliferator-activated receptor-gamma co-activator, the distinct molecular mechanisms and/or molecules influencing obesity or body weight/body mass regulations are not known.

Therefore, the technical problem underlying the present invention was to provide for means and methods for modulating (pathological) metabolic conditions influencing body-weight regulation and/or energy homeostatic circuits. The solution to said technical problem is achieved by providing the embodiments characterized in the claims.

Accordingly, the present invention relates to genes with novel functions in body-weight regulation, energy homeostasis, metabolism, and obesity. The present invention discloses a specific gene involved in the regulation of body-weight, energy homeostasis, metabolism, and obesity, and thus in

disorders related thereto such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, and gallstones. The present invention describes the human genes encoding casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins as being involved in those conditions mentioned above.

The term 'GenBank Accession number' relates to NCBI GenBank database entries (Benson et al, Nucleic Acids Res. 28, 2000, 15-18).

The casein kinase I (CKI) family of protein kinases is a group of highly related, ubiquitously expressed serine/threonine kinases found in all eukaryotic organisms from protozoa to man (Vielhaber and Virshup, 2001, IUBMB Life 51(2):73-78). Recent advances in diverse fields, including developmental biology and chronobiology, have elucidated roles for CKI in regulating critical processes such as Wnt signaling, circadian rhythm, nuclear import, and Alzheimer's disease progression. Casein kinase I is a serine/threonine-specific protein kinase that constitutes most of the kinase activity in eukaryotic cells, where it is mainly localized in the nucleus, cytoplasm, and several membranes. The monomeric enzyme phosphorylates hierarchically a variety of substrates without the involvement of the second messenger in signal transduction.

Casein kinase I, one of the first protein kinases identified biochemically, is known to exist in multiple isoforms in mammals. Three separate members of the CKI gamma subfamily were identified in testis: the isoforms CKI gamma 1, CKI gamma 2, and CKI gamma 3. The proteins are more than 90% identical to each other within the protein kinase domain but only 51-59% identical to other casein kinase I isoforms within this region. Message RNA for CKI gamma 3 was observed in testis, brain, heart, kidney, lung, liver, and muscle whereas CKI gamma 1 and CKI gamma 2 messages were restricted to testis (Zhai et al., 1995, J Biol Chem

270(21):12717-12724). As shown in this invention, Taqman analysis revealed ubiquitous expression of CKI gamma 1 and CKI gamma 2, with strongest expression in testis. The enzymes phosphorylate typical in vitro casein kinase I substrates such as casein, phosvitin, and a synthetic peptide, D4. The known casein kinase I inhibitor CKI-7 also inhibits the CKI gamma's although less effectively than the CKI alpha or CKI delta isoforms. All three CKI gamma's undergo autophosphorylation when incubated with ATP and Mg2+. The YCKI and YCK2 genes in Saccharomyces cerevisiae encode casein kinase I homologs, defects in which lead to aberrant morphology and growth arrest (Zhai et al., supra).

The GABA(A)-receptor-associated protein (GABARAP) is a small 17kDa microtubule associated protein that recognizes and binds the gamma subunit of Type A receptors of gamma-aminobutyric acid (GABA(A)) receptors which plays a central role in the synaptic targeting. GABARAP has also been reported to bind N-ethylmaleimide sensitive factor (NSF), a protein critical for intracellular trafficking events. GABARAP is specifically localized to intracellular membranes, including the Golgi network. The crystal structure of human GABARAP comprises an N-terminal helical subdomain and a ubiquitin-like C-terminal domain (Coyle et al., 2002, Neuron 33(1):63-74). Structure-based mutational analysis demonstrates that the N-terminal subdomain is responsible for tubulin binding while the C-terminal domain contains the binding site for the GABA(A). Coyle et al. (supra) show GABARAP can switch from a monomer to an extended linear polymer form that may function to assemble microtubules during the intracellular trafficking or postsynaptic clustering of GABA(A) receptors. Using the yeast two-hybrid screen, GABARAP has been identified as interactor of ULK1 (Unc-51-like kinase), suggesting an involvement in vesicle transport and axonal elongation in mammalian neurons (Okazaki et al., 2000, Brain Res. Mol. Brain Res. 86:1-12). No function in the regulation of metabolism has been reported for GABARAP or its human homolog.

The dinuclear metalloenzyme methionine aminopeptidases (MAPs) are proteases with important roles in protein processing, especially in proteolysis and peptidolysis (Datta B., 2000, Biochimie. 82(2):95-107). MAPs are involved in the removal of the N-terminal methionine from proteins and peptides (Lowther & Matthews, 2000, Biochim Biophys Acta 1477(1-2):157-167).

Highly homologous MAPs have been identified from various prokaryotic and eukaryotic organisms, for example E. coli, S. typhimurium, P. furiosus, Saccharomyces cerevisiae, Drosophila melanogaster, porcine, mouse, rat, and human. The Drosophila melanogaster gene CG10576 encodes a metallopeptidase family M24 methionyl aminopeptidase (EC:3.4.11.18). A cell cycle-specifically modulated nuclear protein of 38 kDa (termed p38-2G4; PA2G4; ErbB-3 binding protein Ebp1) has been described to be ubiquitously expressed in mouse and human (Radomski & Jost, 1995, Exp Cell Res 220(2):434-445; Lamartine et al., 1997, Cytogenet Cell Genet 78(1):31-35). Substantial progress has recently been made in determining the structures of several members of this family.

The identification of human MAPs as the target of putative anti-cancer drugs reiterates the importance of this family of enzymes. For example, the ErbB-3 binding protein (Ebp1; identical to PA2G4) which is interacting with the juxtamembrane domain of ErbB-3 which is human epidermal growth factor receptor-3 (class I tyrosine kinase receptor) involved in signal transduction pathways that regulate cell growth and differentiation. ErbB-3 has low tyrosine kinase activity, suggesting that it may function more as an adaptor in signaling than as a kinase. The binding of Ebp1 to ErbB-3 inhibits the proliferation and induces the differentiation of human breast cancer cells. The mechanisms of these effects are unknown (see, for example, Lessor et al., 2000, J Cell Physiol 183(3):321-329, Yoo et al., 2000, Br J Cancer 82(3):683-690; Xia et al., 2001, J Cell Physiol 187(2):209-417).

The Drosophila gene cassette Mocs1 encodes for Molybdenum cofactor synthesis-step 1 proteins A, A-B, and B (Mocs1A, Mocs1A-B, and Mocs1B) which are involved in Molybdopterin cofactor biosynthesis. As shown in this invention, Mocs1 is most homologous to the isoforms of human molybdenum cofactor biosynthesis protein 1. Molybdenum is an essential cofactor in many enzymes, but must first be complexed by molybdopterin, whose synthesis requires four enzymatic activities (see, for example, Gray & Nicholls, 2000, RNA 6(7):928-36). The first two enzymes of this pathway are encoded by the MOCS1 locus in humans. A well-conserved novel mRNA splicing phenomenon produces both an apparently bicistronic transcript, as well as a distinct class of monocistronic transcripts. The latter are created by a variety of splicing mechanisms resulting in fusion of the MOCS1A and MOCS1B open reading frames. Therefore, a single bifunctional protein is encoded embodying both MOCS1A and MOCS1B activities. This coexpression profile was observed in vertebrates (including human, mouse, cow, rabbit, opossum, and chicken) and invertebrates (e.g. fruit fly and nematode) spanning at least 700 million years of evolution.

It has been described that Molybdate (Mo) exerts insulinomimetic effects in vitro. Reul et al. (1997, J Endocrinol 155(1):55-64) showed that Mo can improve glucose homeostasis in genetically obese, insulin-resistant ob/ob mice. Oral administration of Mo for 7 weeks did not affect body weight, but decreased the hyperglycaemia of obese mice to the levels of lean (L) (+/+) mice, and reduced the hyperinsulinaemia to one-sixth of pretreatment levels.

Human MoCo deficiency is a fatal disease resulting in severe neurological damage and death in early childhood. Molybdenum cofactor (MoCo) deficiency leads to a combined deficiency of the molybdoenzymes. Effective therapy is not available for this rare disease. Most patients harbor MOCS1 mutations, which prohibit formation of a precursor, or carry

MOCS2 mutations, which abrogate precursor conversion to molybdopterin. A gephyrin gene (GEPH) deletion was identified in a patient with symptoms typical of MoCo deficiency (Reiss et al., 2001, Am J Hum Genet 68(1):208-13). Gephyrin was originally identified as a membrane-associated protein that is essential for the postsynaptic localization of receptors for the neurotransmitters glycine and GABA(A).

Septins are novel GTPase proteins that are broadly distributed in many eukaryotes except plants. The septins are an evolutionary conserved family of proteins that are involved in cytokinesis (the final event of the cell division cycle) and other aspects of cell-surface organization (reviewed in Cooper & Kiehart, 1996; Field & Kellog, 1999). Members of the septin family contain sequences characteristic of the GTPase superfamily of proteins.

For example, in Saccharomyces cerevisiae, the Cdc3, Cdc10, Cdc11, Cdc12 and Shs1/Sep7 septins assemble as a ring that marks the cytokinetic plane throughout the budding cycle (see, for example, Sidd et al, 2001, Microbiology 147(Pt 6):1437-50). This structure participates in different aspects of morphogenesis, such as selection of cell polarity, localization of chitin synthesis, the switch from hyperpolar to isotropic bud growth after bud emergence and the spatial regulation of septation. The septin cytoskeleton assembles at the pre-bud site before bud emergence, remains there during bud growth and duplicates at late mitosis eventually disappearing after cell separation. The high degree of conservation, ubiquitous expression and proven role in cytokinesis suggests septins are certain to be important players in regulating cell architecture and function (see, for example, Field et al., 1996, J Cell Biol 133(3):605-616).

For example, the Drosophila gene peanut (pnut) encodes a septin homolog, microtubule binding protein involved in cytokinesis (see, for example, Neufeld and Rubin, 1994, Cell 77(3):371-379). Pnut protein is localized to

the cleavage furrow of dividing cells during cytokinesis and to the intercellular bridge connecting postmitotic daughter cells. In addition to its role in cytokinesis, pnut displays genetic interactions with seven in absentia (sina), a gene required for neuronal fate determination in the compound eye and involved in ubiquitin-dependent protein degradation. The amino acid sequence of the Drosophila gene pnut is highly homologous to that of Saccharomyces cerevisiae CDC3, CDC10, CDC11, CDC12, Candida alicancs (CaCDC10), the Drosophila genes Sep1, Sep2, and the mammalian genes BH5, cell cycle division 10 (CDC10), septin Nedd5, Diff6, septin 2 (Sep2), and septin 3 (Sep3), which are implicated in cytokinesis and cell polarity (Xiong et al., 1999, Mech Dev 86(1-2):183-191).

Enzymes of the glycolytic pathway convert the sugar glucose to pyruvate while simultaneously producing ATP. The pathway also provides building blocks for the synthesis of cellular components such as long-chain fatty acids. After glycolysis, pyruvate is converted to acetyl-Coenzyme A, which enters the citric acid cycle. Glycolytic enzymes include hexokinase, phosphoglucose3-isomerase, phosphofructokinase, aldolase, triose, phosphate isomerase, glyceraldehyde, 3-phosphatedehydrogenase, phosphoglycerate kinase, phosphoglyceromutase, enolase, and pyruvate kinase. Of these, phosphofructokinase, hexokinase, and pyruvate kinase are important in regulating the rate of glycolysis.

Carbohydrates mediate their conversion to triglycerides in the liver by promoting both rapid posttranslational activation of rate-limiting glycolytic and lipogenic enzymes and transcriptional induction of the genes encoding many of these same enzymes. A transcription factor has been described that recognizes the carbohydrate response element (ChRE) within the promoter of the L-type pyruvate kinase (LPK) gene. The DNA-binding activity of this ChRE-binding protein in rat livers is specifically induced by a high carbohydrate diet. It was suggested that the ChRE-binding protein

may contribute to the imbalance between nutrient utilization and storage characteristic of obesity (Yamashita et al., 2001, Proc Natl Acad Sci U S A 31;98(16):9116-21

Obese (fa/fa) Zucker rat shows altered thermogenesis and changes in both lipid and carbohydrate metabolism (see, for example, Sanchez-Gutierrez, 2000, Arch Biochem Biophys 373(1):249-54; Perez et al., 1998, Int J Obes Relat Metab Disord 22(7):667-72). The activities of glucokinase and L-pyruvate kinase increased in fed obese (fa/fa) rats compared with fed lean (fa/-) animals, but decreased during starvation. The mRNA levels of glycolytic enzymes such as glucokinase and L-pyruvate kinase in fed obese rats were higher than in fed lean animals. During starvation, they decreased in lean and obese rats. The stimulation of gluconeogenesis by epinephrine was accompanied by an inactivation of both pyruvate kinase and 6-phosphofructo 2-kinase in rat hepatocytes.

Pyruvate kinase is a key enzyme in glycogen metabolism. Mammalian pyruvate kinases of different tissues are distinct, their characteristics being related to tissue metabolic requirements (for example see Bigley et al., 1974, Enzyme 17(5):297-306). Pyruvate kinase is also known as ATP:pyruvate phosphotransferase (EC 2.7.1.40). At least 3 molecular forms with pyruvate kinase activity are known (Bigley et al., 1968, Enzym. Biol. Clin. 9: 10-20). The form that is deficient in a type of hemolytic anemia is the red cell variety, PK1. PK2 is found in kidney. PK3 is found in leukocytes, muscle, platelets, and brain but not in red cells or kidney. PK3 is a tetrameric protein and all subunits are alike. The enzyme is insensitive to fructose-1,6-diphosphate. Tsutsumi et al. (1988, Genomics 2(1):86-9) showed that pyruvate kinase occurs in 4 isozymic forms (L, R, M1, M2).

The Drosophila gene CG9429 (Crc, calreticulin) encodes for a putative calcium binding protein (chaperone) which is a component of the endoplasmic reticulum in Drosophila. Intrapro analysis of this gene reveals

an endoplasmic reticulum targeting sequence, calreticulum family domains, and aspartic acid-rich regions. As shown in this invention, the Drosophila Crc is most homologous to human calreticulin (other names: CRP55, calregulin, HACBP, ERP60, 52 kDa Ribonucleoprotein autoantigen RO/SS-A, sicca syndrome antigen A or autoantigen Ro; for example, GenBank Accession Number NM_004343 and XM_032030 (identical proteins)) and to mouse calreticulin (GenBank Accession Numbers AAH03453.1, AAH03453, and BC003453). Calreticulin is a highly conserved, multifunctional protein that acts as a major calcium-binding protein, most abundant in the lumen of the endoplasmic and sarcoplasmic reticulum. The protein has well-recognized physiological roles in the ER as a molecular chaperone and Ca(2+)-signalling molecule. Calreticulin has also been found in other membrane-bound organelles, at the cell surface and in the extracellular environment, where it has recently been shown to exert a number of physiological and pathological effects, see, for example, review by Johnson et al, 2001, Trends Cell Biol 11(3):122-9. In addition to the calcium-binding functions and molecular chaperone function, calreticulin has been characterized as an extracellular lectin, an intracellular mediator of integrin function, an inhibitor of steroid hormone-regulated gene expression and a C1q-binding protein (see, for example, review by Coppolino et al., 1998, Int J Biochem Cell Biol 30(5):553-8). Calreticulin binds to antibodies in certain sera of systemic lupus and Sjogren patients which contain anti-Ro/SSA antibodies. Increased autoantibody titer against human calreticulin is found in infants with complete congenital heart block of both the IgG and IgM classes.

So far, it has not been described that casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins are involved in the regulation of energy homeostasis and body-weight regulation and related disorders, and thus, no functions in metabolic diseases and other diseases as listed above have been discussed. In this invention we demonstrate that the correct gene dose of casein kinase 1

gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins is essential for maintenance of energy homeostasis. A genetic screen was used to identify that mutation of genes encoding casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin homologous proteins cause obesity, reflected by a significant increase of triglyceride content, the major energy storage substance.

The function of calreticulin and casein kinase 1 gamma in metabolic disorders is further validated by data obtained from additional screens. For example, an additional screen using Drosophila mutants with modifications of the eye phenotype identified a modification of UCP activity by calreticulin, thereby leading to an altered mitochondrial activity. An additional screen using Drosophila mutants with modifications of the eye phenotype identified an interaction of casein kinase 1 gamma with adipose, a protein regulating, causing or contributing to obesity. These findings suggest the presence of similar activities of these described homologous proteins in humans that provides insight into diagnosis, treatment, and prognosis of metabolic disorders.

Polynucleotides encoding proteins with homologies to casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin are suitable to investigate diseases and disorders as described above. Further new compositions useful in diagnosis, treatment, and prognosis of diseases and disorders as described above are provided.

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular methodology, protocols, cell lines, vectors, and reagents described as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention that will be limited only by the appended claims. Unless defined otherwise, all technical and

scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods, devices, and materials are now described. All publications mentioned herein are incorporated herein by reference for the purpose of describing and disclosing the cell lines, vectors, and methodologies that are reported in the publications which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure.

The present invention discloses that casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins are regulating the energy homeostasis and fat metabolism especially the metabolism and storage of triglycerides, and polynucleotides, which identify and encode the proteins disclosed in this invention. The invention also relates to vectors, host cells, antibodies, and recombinant methods for producing the polypeptides and polynucleotides of the invention. The invention also relates to the use of these sequences in the diagnosis, study, prevention, and treatment of diseases and disorders, for example, but not limited to, metabolic diseases such as obesity as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, and gallstones.

Casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, and calreticulin homologous proteins and nucleic acid molecules coding therefore are obtainable from insect or vertebrate species, e.g. mammals or birds. Particularly preferred are homologous nucleic acids, particularly nucleic acids encoding a human casein kinase 1, gamma 1, human casein kinase 1, gamma 2, human casein kinase 1, gamma 3, human GABARAP,

human GABARAP like 1, human GABARAP like 2, human GABARAP like 3, human proliferation-associated 2G4 protein, the human MOCS1 isoforms, human CDC10, human pyruvate kinase, muscle, human pyruvate kinase, liver and RBC, human calreticulin, and/or human calreticulin 2.

The invention particularly relates to a nucleic acid molecule encoding a polypeptide contributing to regulating the energy homeostasis and the metabolism of triglycerides, wherein said nucleic acid molecule comprises

(a) the nucleotide sequence of (i) Drosophila gilgamesh (gish), human casein kinase 1, gamma 1 (SEQ ID NO: 1), human casein kinase 1, gamma 2 (SEQ ID NO: 3), human casein kinase 1, gamma 3 (SEQ ID NO: 5), (ii) Drosophila Gadfly Accession Number CG1534 (synonym for Gadfly Accession Number CG32672 and Genbank Accession Number. NM_167245), human GABARAP (SEQ ID NO: 7), human GABARAP like 1 (SEQ ID NO: 9), human GABARAP like 2 (SEQ ID NO: 11), human GABARAP like 3 (SEQ ID NO: 13), (iii) Drosophila Gadfly Accession Number CG10576, human PA2G4 (SEQ ID NO: 15), (iv) Drosophila Mocs1, human MOCSA (SEQ ID NO: 17), human MOCS1 isoform 1 (SEQ ID NO: 19), human MOCS1 isoform 2 (SEQ ID NO: 21), human MOCS1 isoform 3 (SEQ ID NO: 23), (v) Drosophila peanut (pnut), human CDC10 (SEQ ID NO: 25), (vi) Drosophila Gadfly Accession Number CG7069, human pyruvate kinase, muscle (SEQ ID NO: 27), human pyruvate kinase, liver and RBC (SEQ ID NO: 30), (vii) Drosophila calreticulin (Crc), human calreticulin (SEQ ID NO: 32), human calreticulin 2 (SEQ ID NO:34), and/or a sequence complementary thereto,

(b) a nucleotide sequence which hybridizes at 50°C in a solution containing 1 x SSC and 0.1% SDS to a sequence of (a),

(c) a sequence corresponding to the sequences of (a) or (b) within the degeneration of the genetic code,

(d) a sequence which encodes a polypeptide which is at least 85%, preferably at least 90%, more preferably at least 95%, more

preferably at least 98% and up to 99,6% identical to the amino acid sequences of human casein kinase 1, gamma 1 (SEQ ID NO: 2), human casein kinase 1, gamma 2 (SEQ ID NO: 4), human casein kinase 1, gamma 3 (SEQ ID NO: 6), human GABARAP (SEQ ID NO: 8), human GABARAP like 1 (SEQ ID NO: 10), human GABARAP like 2 (SEQ ID NO: 12), human GABARAP like 3 (SEQ ID NO: 14), human PA2G4 (SEQ ID NO: 16), human MOCSA (SEQ ID NO: 18), human MOCS1 isoform 1 (SEQ ID NO: 20), human MOCS1 isoform 2 (SEQ ID NO: 22), human MOCS1 isoform 3 (SEQ ID NO: 24), human CDC10 (SEQ ID NO: 26), human pyruvate kinase, muscle, isozyme M1 (SEQ ID NO: 28), human pyruvate kinase, muscle, isozyme M2 (SEQ ID NO: 29), human pyruvate kinase, liver and RBC (SEQ ID NO: 31), human calreticulin (SEQ ID NO: 33), and/or human calreticulin 2 (SEQ ID NO:35),

(e) a sequence which differs from the nucleic acid molecule of (a) to (d) by mutation and wherein said mutation causes an alteration, deletion, duplication and/or premature stop in the encoded polypeptide or

(f) a partial sequence of any of the nucleotide sequences of (a) to (e) having a length of at least 15 bases, preferably at least 20 bases, more preferably at least 25 bases and most preferably at least 50 bases.

The invention is based on the discovery that casein kinase 1 gamma (CSNK1G), GABA(A) receptor-associated protein (GABARAP), proliferation-associated 2G4 protein, 38kDa (PA2G4, also referred to as methionyl aminopeptidase homologous protein), molybdenum cofactor synthesis-step 1 protein (MOCS1), cell division cycle 10 protein homolog (CDC10, also referred to as septin 7), pyruvate kinase (PK), calreticulin (CALR) or homologous proteins (herein referred to as casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin) and the polynucleotides encoding these, are involved in the regulation of

triglyceride storage and therefore energy homeostasis. The invention describes the use of these polypeptides or fragments thereof, polynucleotides or fragments thereof and effectors (receptors) of these molecules, e.g. antibodies, biologically active nucleic acids, such as antisense molecules, RNAi molecules or ribozymes, aptamers, peptides or low-molecular weight organic compounds recognizing said polynucleotides or polypeptides for the diagnosis, study, prevention, or treatment of diseases and disorders related thereto, including metabolic diseases such as obesity as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, and gallstones.

Accordingly, the present invention relates to genes with novel functions in body-weight regulation, energy homeostasis, metabolism, and obesity. To find genes with novel functions in energy homeostasis, metabolism, and obesity, a functional genetic screen was performed with the model organism Drosophila melanogaster (Meigen). The ability to manipulate and screen the genomes of model organisms such as the fly Drosophila melanogaster provides a powerful tool to analyze biological and biochemical processes that have direct relevance to more complex vertebrate organisms due to significant evolutionary conservation of genes, cellular processes, and pathways (see, for example, Adams M. D. et al., (2000) Science 287: 2185-2195). Identification of novel gene functions in model organisms can directly contribute to the elucidation of correlative pathways in mammals (humans) and of the methods of modulating them. A correlation between a pathology model (such as changes in triglyceride levels as indication for metabolic syndrome including obesity) and the modified expression of a fly gene can identify the association of the human ortholog with the particular human disease. One resource for screening was a proprietary Drosophila melanogaster stock collection of EP-lines. Additionally, the publicly available EP-collection was screened. The P-vector of both collections has Gal4-UAS-binding sites fused to a basal

promoter that can transcribe adjacent genomic Drosophila sequences upon binding of Gal4 to UAS-sites. This enables the EP-line collection for overexpression of endogenous flanking gene sequences. In addition, without activation of the UAS-sites, integration of the EP-element into the gene is likely to cause a reduction of gene activity, and allows determining its function by evaluating the loss-of-function phenotype.

Triglycerides are the most efficient storage for energy in cells. Obese people mainly show a significant increase in the content of triglycerides. In order to isolate genes with a function in energy homeostasis, several thousand proprietary and publicly available EP-lines were tested for their triglyceride content after a prolonged feeding period (see Examples for more detail). Lines with significantly changed triglyceride content were selected as positive candidates for further analysis.

In this invention, the content of triglycerides of a pool of flies with the same genotype was analyzed after feeding for six days using a triglyceride assay, as, for example, but not for limiting the scope of the invention, is described below in the examples section. Male flies homozygous for the integration of vectors for Drosophila lines HD-EP(3)37409, EP(3)3271, EP(3)3688, EP(2)2036, EP(3)3224, EP(3)3321, EP(3)0834, and EP(3)0979, and hemizygous for the integration of vectors for Drosophila line PX6298.1, were analyzed in assays measuring the triglyceride contents of these flies, illustrated in more detail in the EXAMPLES section. The results of the triglyceride content analysis are shown in FIGURES 1, 6, 11, 16, 21, 26, and 30.

Adipose (adp) is a protein that has been described as regulating, causing or contributing to obesity in an animal or human (see WO 01/96371). Transgenic flies over-expressing the adipose gene in the developing Drosophila eye were generated and analyzed for modifications of the eye phenotype (for example, an enhancement or a suppression of the

phenotype). Mutations changing the eye phenotype affect genes that modify the activity of adipose. Fly line HD-EP(3)37409 was found to be an enhancer of the described eye phenotype. This result is strongly suggesting an interaction of gilgamesh gene with adipose since the integration of HD-EP(3)37409 was found to be located at the gilgamesh locus. This is supporting the function of gilgamesh and homologous proteins in the regulation of the energy homeostasis.

An additional screen using Drosophila mutants with modifications of the eye phenotype identified a modification of UCP activity by calreticulin, thereby leading to an altered mitochondrial activity.

Genomic DNA sequences were isolated that are localized to the EP vector (herein HD-EP(3)37409, PX6298.1, EP(3)3271, EP(3)3688, EP(2)2036, EP(3)3224, EP(3)3321, EP(3)0834, and EP(3)0979) integration. Using those isolated genomic sequences public databases like Berkeley Drosophila Genome Project (GadFly; see also FlyBase (1999) Nucleic Acids Research 27:85-88) were screened thereby identifying the integration site of the vectors, and the corresponding gene, described in more detail in the EXAMPLES section. The molecular organization of the gene is shown in FIGURES 2, 7, 12, 17, 22, 27, and 31.

The Drosophila genes and proteins encoded thereby with functions in the regulation of triglyceride metabolism were further analysed using the BLAST algorithm searching in publicly available sequence databases and mammalian homologs were identified (see FIGURES 3, 4, 8, 9, 13, 14, 18, 19, 23, 24, 28, 29, 32, and 33).

The function of the mammalian homologs in energy homeostasis was further validated in this invention by analyzing the expression of the transcripts in different tissues and by analyzing the role in adipocyte differentiation. Expression profiling studies (see Examples for more detail)

confirm the particular relevance of the protein(s) of the invention as regulators of enery metabolism in mammals. Further, we show that the proteins of the invention are regulated by fasting and by genetically induced obesity. In this invention, we used mouse models of insulin resistance and/or diabetes, such as mice carrying gene knockouts in the leptin pathway (for example, ob (leptin) or db (leptin receptor) mice) to study the expression of the protein of the invention. Such mice develop typical symptoms of diabetes, show hepatic lipid accumulation and frequently have increased plasma lipid levels (see Bruning et al, 1998, Mol. Cell. 2:449-569).

The invention also encompasses polynucleotides that encode the proteins of the invention and homologous proteins. Accordingly, any nucleic acid sequence, which encodes the amino acid sequences of the proteins of the invention and homologous proteins, can be used to generate recombinant molecules that express the proteins of the invention and homologous proteins. In a particular embodiment, the invention encompasses a nucleic acid encoding (i) Drosophila gilgamesh (gish), human casein kinase 1, gamma 1, human casein kinase 1, gamma 2, human casein kinase 1, gamma 3, (ii) Drosophila Gadfly Accession Number CG1534 (synonym for Gadfly Accession Number CG32672 and Genbank Accession Number. NM_167245), human GABARAP, human GABARAP like 1, human GABARAP like, human GABARAP like 3, (iii) Drosophila Gadfly Accession Number CG10576, human PA2G4, (iv) Drosophila Mocs1, human MOCSA, human MOCS1 isoform 1, human MOCS1 isoform 2, human MOCS1 isoform 3, (v) Drosophila peanut (pnut), human CDC10, (vi) Drosophila Gadfly Accession Number CG7069, human pyruvate kinase, muscle, human pyruvat kinase, liver and RBC, (vii) Drosophila calreticulin (Crc), human calreticulin, or human calreticulin 2; referred to herein as the proteins of the invention. It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of nucleotide sequences encoding casein kinase 1 gamma, GABARAP,

PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins, some bearing minimal homology to the nucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of nucleotide sequence that could be made by selecting combinations based on possible codon choices.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed nucleotide sequences, and in particular, those of the polynucleotide encoding (i) Drosophila gilgamesh (gish), human casein kinase 1, gamma 1, human casein kinase 1, gamma 2, human casein kinase 1, gamma 3, (ii) Drosophila Gadfly Accession Number CG1534 (synonym for Gadfly Accession Number CG32672 and Genbank Accession Number. NM_167245), human GABARAP, human GABARAP like 1, human GABARAP like, human GABARAP like 3, (iii) Drosophila Gadfly Accession Number CG10576, human PA2G4, (iv) Drosophila Mocs1, human MOCSA, human MOCS1 isoform 1, human MOCS1 isoform 2, human MOCS1 isoform 3, (v) Drosophila peanut (pnut), human CDC10, (vi) Drosophila Gadfly Accession Number CG7069, human pyruvate kinase, muscle, human pyruvate kinase, liver and RBC, (vii) Drosophila calreticulin (Crc), human calreticulin, or human calreticulin 2, under various conditions of stringency. Hybridization conditions are based on the melting temperature (Tm) of the nucleic acid binding complex or probe, as taught in Wahl, G. M. and S. L. Berger (1987: Methods Enzymol. 152:399-407) and Kimmel, A. R. (1987; Methods Enzymol. 152:507-511), and may be used at a defined stringency. Preferably, hybridization under stringent conditions means that after washing for 1 h with 1 x SSC and 0.1% SDS at 50°C, preferably at 55°C, more preferably at 62°C and most preferably at 68°C, particularly for 1 h in 0.2 x SSC and 0.1% SDS at 50°C, preferably at 55°C, more preferably at 62°C and most preferably at 68°C, a positive hybridization signal is observed. Altered nucleic acid sequences encoding casein kinase 1 gamma, GABARAP, PA2G4, MOCS1,

CDC10, PK, calreticulin, or homologous proteins which are encompassed by the invention include deletions, insertions, or substitutions of different nucleotides resulting in a polynucleotide that encodes the same or a functionally equivalent casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins.

The encoded proteins may also contain deletions, insertions, or substitutions of amino acid residues, which produce a silent change and result in functionally equivalent casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues as long as the biological activity of casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins is retained. Furthermore, the invention relates to peptide fragments of the proteins or derivatives thereof such as cyclic peptides, retro-inverso peptides mimetics having a length of at least 4, preferably at least 6 and up to 50 amino acids.

Also included within the scope of the present invention are alleles of the genes encoding casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins. As used herein, an "allele" or "allelic sequence" is an alternative form of the gene, which may result from at least one mutation in the nucleic acid sequence. Alleles may result in altered mRNAs or polypeptides whose structures or function may or may not be altered. Any given gene may have none, one, or many allelic forms. Common mutational changes, which give rise to alleles, are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

The nucleic acid sequences encoding casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins may be extended utilizing a partial nucleotide sequence and employing various methods known in the art to detect upstream sequences such as promoters and regulatory elements. For example, one method which may be employed, "restriction-site" PCR, uses universal primers to retrieve unknown sequence adjacent to a known locus (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). Inverse PCR may also be used to amplify or extend sequences using divergent primers based on a known region (Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186).

Another method which may be used is capture PCR which involves PCR amplification of DNA fragments adjacent to a known sequence in human and yeast artificial chromosome DNA (Lagerstrom, M. et al. (PCR Methods Applic. 1:111-119). Another method which may be used to retrieve unknown sequences is that of Parker, J. D. et al. (1991; Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries to walk in genomic DNA (Clontech, Palo Alto, Calif.). This process avoids the need to screen libraries and is useful in finding intron/exon junctions.

In order to express a biologically active casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous protein, the nucleotide sequences encoding casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins functional equivalents, may be inserted into appropriate expression vectors, i.e., a vector, which contains the necessary elements for the transcription and translation of the inserted coding sequence. Methods, which are well known to those skilled in the art, may be used to construct expression vectors containing sequences encoding casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins and appropriate transcriptional and translational control elements.

Regulatory elements include for example a promoter, an initiation codon, a stop codon, a mRNA stability regulatory element, and a polyadenylation signal. Expression of a polynucleotide can be assured by (i) constitutive promoters such as the Cytomegalovirus (CMV) promoter/enhancer region, (ii) tissue specific promoters such as the insulin promoter (see, Soria et al., 2000, Diabetes 49:157), SOX2 gene promoter (see Li et al., (1998) Curr. Biol. 8:971-974), Msi-1 promoter (see Sakakibara et al., (1997) J. Neuroscience 17:8300-8312), alpha-cardia myosin heavy chain promoter or human atrial natriuretic factor promoter (Klug et al., (1996) J. clin. Invest 98:216-224; Wu et al., (1989) J. Biol. Chem. 264:6472-6479) or (iii) inducible promoters such as the tetracycline inducible system. Expression vectors can also contain a selection agent of marker gene that confers antibiotic resistance such as the neomycin, hygromycin or puromycin resistance genes. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. Such techniques are described in Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview, N.Y., and Ausubel, F. M. et al. (1989) Current Protocols in Molecular Biology, John Wiley & Sons, New York, N.Y.

A variety of expression vector/host systems may be utilized to contain and express sequences encoding casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins. These include, but are not limited to, micro-organisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (e.g., baculovirus); plant cell systems transformed with virus expression vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or PBR322 plasmids); or animal cell systems. The "control elements" or "regulatory sequences" are those non-translated regions of the vector-enhancers, promoters, 5′ and 3′ untranslated regions

which interact with host cellular proteins to carry out transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used.

The presence of polynucleotide sequences encoding casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins can be detected by DNA-DNA or DNA-RNA hybridization and/or amplification using probes or portions or fragments of polynucleotides encoding casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins. Nucleic acid amplification based assays involve the use of oligonucleotides or oligomers based on the sequences encoding casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins to detect transformants containing DNA or RNA encoding casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins. As used herein "oligonucleotides" or "oligomers" refer to a nucleic acid sequence of at least about 10 nucleotides and as many as about 60 nucleotides, preferably about 15 to 30 nucleotides, and more preferably about 20-25 nucleotides, which can be used as a probe or amplimer.

The presence of proteins of the invention in a sample can be determined by immunological methods or activity measurement. A variety of protocols for detecting and measuring the expression of casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins, using either polyclonal or monoclonal antibodies specific for the protein are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on

casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins is preferred, but a competitive binding assay may be employed. These and other assays are described, among other places, in Hampton, R. et al. (1990; Serological Methods, a Laboratory Manual, APS Press, St Paul, Minn.) and Maddox, D. E. et al. (1983; J. Exp. Med. 158:1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins include oligo-labeling, nick translation, end-labeling or PCR amplification using a labeled nucleotide, or enzymatic synthesis.

Alternatively, the sequences encoding casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins, or any portions thereof may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits (Pharmacia & Upjohn, (Kalamazoo, Mich.); Promega (Madison Wis.); and U.S. Biochemical Corp., (Cleveland, Ohio).

Suitable reporter molecules or labels, which may be used, include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, co-factors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or

homologous proteins may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a recombinant cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins may be designed to contain signal sequences, which direct secretion of casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins through a prokaryotic or eukaryotic cell membrane. Other recombinant constructions may be used to join sequences encoding casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins to nucleotide sequence encoding a polypeptide domain, which will facilitate purification of soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAG extension/affinity purification system (Immunex Corp., Seattle, Wash.) The inclusion of cleavable linker sequences such as those specific for Factor XA or Enterokinase (Invitrogen, San Diego, Calif.) between the purification domains and casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins may be used to facilitate purification.

The nucleic acids encoding the proteins of the invention can be used to generate transgenic animal or site specific gene modifications in cell lines. Transgenic animals may be made through homologous recombination, where the normal locus of the genes encoding the proteins of the invention is altered. Alternatively, a nucleic acid construct is randomly integrated into the genome. Vectors for stable integration include plasmids, retroviruses and other animal viruses, YACs, and the like. The modified cells or animal

are useful in the study of the function and regulation of the proteins of the invention. For example, a series of small deletions and/or substitutions may be made in the genes that encode the proteins of the invention to determine the role of particular domains of the protein, functions in pancreatic differentiation, etc.

Specific constructs of interest include anti-sense molecules, which will block the expression of the proteins of the invention, or expression of dominant negative mutations. A detectable marker, such as for example lac-Z, may be introduced in the locus of the genes of the invention, where upregulation of expression of the genes of the invention will result in an easily detected change in phenotype.

One may also provide for expression of the genes of the invention or variants thereof in cells or tissues where it is not normally expressed or at abnormal times of development. In addition, by providing expression of the proteins of the invention in cells in which they are not normally produced, one can induce changes in cell behavior.

DNA constructs for homologous recombination will comprise at least portions of the genes of the invention with the desired genetic modification, and will include regions of homology to the target locus. DNA constructs for random integration need not include regions of homology to mediate recombination. Conveniently, markers for positive and/or negative selection are included. Methods for generating cells having targeted gene modifications through homologous recombination are known in the art. For embryonic stem (ES) cells, an ES cell line may be employed, or embryonic cells may be obtained freshly from a host, e.g. mouse, rat, guinea pig etc. Such cells are grown on an appropriate fibroblast-feeder layer or grown in presence of leukemia inhibiting factor (LIF).

When ES or embryonic cells or somatic pluripotent stem cells have been transformed, they may be used to produce transgenic animals. After transformation, the cells are plated onto a feeder layer in an appropriate medium. Cells containing the construct may be detected by employing a selective medium. After sufficient time for colonies to grow, they are picked and analyzed for the occurrence of homologous recombination or integration of the construct. Those colonies that are positive may then be used for embryo manipulation and blastocyst injection. Blastocysts are obtained from 4 to 6 week old superovulated females. The ES cells are trypsinized, and the modified cells are injected into the blastocoel of the blastocyst. After injection, the blastocysts are returned to each uterine horn of pseudopregnant females. Females are then allowed to go to term and the resulting offspring screened for the construct. By providing for a different phenotype of the blastocyst and the genetically modified cells, chimeric progeny can be readily detected. The chimeric animals are screened for the presence of the modified gene and males and females having the modification are mated to produce homozygous progeny. If the gene alterations cause lethality at some point in development, tissues or organs can be maintained as allogenic or congenic grafts or transplants, or in vitro culture. The transgenic animals may be any non-human mammal, such as laboratory animal, domestic animals, etc. The transgenic animals may be used in functional studies, drug screening, etc.

Diagnostics and Therapeutics

The data disclosed in this invention show that the nucleic acids and proteins of the invention and effector molecules thereof are useful in diagnostic and therapeutic applications implicated, for example but not limited to, in metabolic disorders such as obesity as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia,

osteoarthritis, and gallstones. Hence, diagnostic and therapeutic uses for the casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin nucleic acids and proteins, or homologous proteins of the invention are, for example but not limited to, the following: (i) protein therapeutic, (ii) small molecule drug target, (iii) antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) diagnostic and/or prognostic marker, (v) gene therapy (gene delivery/gene ablation), (vi) research tools, and (vii) tissue regeneration in vitro and in vivo (regeneration for all these tissues and cell types composing these tissues and cell types derived from these tissues).

The nucleic acids and proteins of the invention are useful in diagnostic and therapeutic applications implicated in various applications as described below. For example, but not limited to, cDNAs encoding the casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin proteins of the invention and particularly their human homologues may be useful in gene therapy, and the casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin proteins of the invention and particularly their human homologues may be useful when administered to a subject in need thereof. By way of non-limiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from, for example, but not limited to, in metabolic disorders as described above.

The nucleic acid encoding the casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin proteins of the invention, or homologous proteins, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acids or the proteins are to be assessed. These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods.

For example, in one aspect, antibodies which are specific for casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins may be used directly as an antagonist, or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins. The antibodies may be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimerical, single chain, Fab fragments, and fragments produced by a Fab expression library. Neutralising antibodies, (i.e., those which inhibit dimer formation) are especially preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others, may be immunized by injection with casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminium hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, and dinitrophenol. Among adjuvants used in human, BCG (Bacille Calmette-Guerin) and Corynebacterium parvum are especially preferable. It is preferred that the peptides, fragments, or oligopeptides used to induce antibodies to casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins have an amino acid sequence consisting of at least five amino acids, and more preferably at least 10 amino acids.

Monoclonal antibodies to casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins may be prepared using any technique that provides for the production of antibody molecules

by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Köhler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R. J. et al. Proc. Natl. Acad. Sci. 80:2026-2030; Cole, S. P. et al. (1984) Mol. Cell Biol. 62:109-120).

In addition, techniques developed for the production of "chimeric antibodies", the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity can be used (Morrison, S. L. et al. (1984) Proc. Natl. Acad. Sci. 81:6851-6855; Neuberger, M. S. et al (1984) Nature 312:604-608; Takeda, S. et al. (1985) Nature 314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins - and -specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries (Burton, D. R. (1991) Proc. Natl. Acad. Sci. 88:11120-3). Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening recombinant immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature (Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299).

Antibody fragments, which contain specific binding sites for casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins, may also be generated. For example, such fragments include, but are not limited to, the $F(ab')_2$ fragments which can be produced by Pepsin digestion of the antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of

$F(ab')_2$ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse, W. D. et al. (1989) Science 254:1275-1281).

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding and immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins and its specific antibody. A two-site, monoclonal-based immunoassay utilising monoclonal antibodies reactive to two non-interfering casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous protein epitopes are preferred, but a competitive binding assay may also be employed (Maddox, supra).

In another embodiment of the invention, the polynucleotides encoding casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins, or any fragment thereof, or nucleic acid effector molecules such as antisense molecules, aptamers, RNAi molecules or ribozymes may be used for therapeutic purposes. In one aspect, aptamers, i.e. nucleic acid molecules, which are capable of binding to a protein of the invention and modulating its activity, may be generated by a screening and selection procedure involving the use of combinatorial nucleic acid libraries.

In a further aspect, antisense molecules, may be used for therapeutic purposes. In one aspect, antisense to the polynucleotide encoding casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins may be used in situations in which it would be

desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins. Thus, antisense molecules may be used to modulate casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or and homologous protein activity, or to achieve regulation of gene function. Such technology is now well know in the art, and sense or antisense oligomers or larger fragments, can be designed from various locations along the coding or control regions of sequences encoding casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins. Expression vectors derived from retroviruses, adenovirus, herpes or vaccinia viruses, or from various bacterial plasmids may be used for delivery of nucleotide sequences to the targeted organ, tissue or cell population. Methods, which are well known to those skilled in the art, can be used to construct recombinant vectors, which will express antisense molecules complementary to the polynucleotides of the genes encoding casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins. These techniques are described both in Sambrook et al. (supra) and in Ausubel et al. (supra). Genes encoding casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins can be turned off by transforming a cell or tissue with expression vectors which express high levels of polynucleotide or fragment thereof which encodes casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector and even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing antisense molecules, DNA, RNA, or nucleic acid analogues such as PNA, to the control regions of the genes encoding casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins, i.e., the promoters, enhancers, and introns. Oligonucleotides derived from the transcription initiation site, e.g., between positions -10 and +10 from the start site, are preferred. Similarly, inhibition can be achieved using "triple helix" base-pairing methodology. Triple helix pairing is useful because it cause inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J. E. et al. (1994) In; Huber, B. E. and B. I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, N.Y.). The antisense molecules may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. Examples, which may be used, include engineered hammerhead motif ribozyme molecules that can be specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins. Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites which include the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides corresponding to the region of the target gene containing the cleavage site may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by

testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Nucleic acid effector molecules such as antisense molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins. Such DNA sequences may be incorporated into a variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize antisense RNA constitutively or inducibly can be introduced into cell lines, cells, or tissues. RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of non-traditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection and by liposome injections may be achieved using methods, which are well known in the art. Any of the therapeutic methods described above may be applied to any suitable subject including, for

example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

An additional embodiment of the invention relates to the administration of a pharmaceutical composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins, antibodies to casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins, mimetics, agonists, antagonists, or inhibitors of casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins. The compositions may be administered alone or in combination with at least one other agent, such as stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier, including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs or hormones. The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries, which facilitate processing of the active compounds into preparations which, can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, Pa.).

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes. After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art. For any compounds, the therapeutically effective does can be estimated initially either in cell culture assays, e.g., of preadipocyte cell lines, or in animal models, usually mice, rabbits, dogs, or pigs. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans. A therapeutically effective dose refers to that amount of active ingredient, for example casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins or fragments thereof, or antibodies, which is effective against a specific condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population). The dose ratio between therapeutic and toxic effects is the therapeutic index, and it can be expressed as the ratio, LD50/ED50. Pharmaceutical compositions, which exhibit large therapeutic indices, are preferred. The data obtained from cell culture assays and animal studies is used in formulating a range of dosage for human use. The dosage contained in

such compositions is preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage varies within this range depending upon the dosage being employed, the sensitivity of the patient, and the route of administration. The exact dosage will be determined by the practitioner, in light of factors related to the subject that requires treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors, which may be taken into account, include the severity of the disease state, general health of the subject, age, weight, and gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation. Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

In another embodiment, antibodies which specifically bind casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins may be used for the diagnosis of conditions or diseases characterized by or associated with over- or underexpression of casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins, or in assays to monitor patients being treated with casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins, agonists, antagonists or inhibitors. The antibodies useful for diagnostic purposes may be prepared in the same manner as those described above for therapeutics. Diagnostic assays for

casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins include methods, which utilize the antibody and a label to detect casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins in human body fluids or extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by joining them, either covalently or non-covalently, with a reporter molecule. A wide variety of reporter molecules, which are known in the art may be used several of which are described above.

A variety of protocols including ELISA, RIA, and FACS for measuring casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins are known in the art and provide a basis for diagnosing altered or abnormal levels of casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous protein expression. Normal or standard values for casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous protein expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with antibody to casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins under conditions suitable for complex formation. The amount of standard complex formation may be quantified by various methods, but preferably by photometric means. Quantities of casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins expressed in control and disease, samples, e.g. from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins may be used for diagnostic purposes.

The polynucleotides, which may be used, include oligonucleotide sequences, antisense RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantitate gene expression in biopsied tissues in which expression of casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins may be correlated with disease. The diagnostic assay may be used to distinguish between absence, presence, and excess expression of casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins, and to monitor regulation of casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous protein levels during therapeutic intervention.

In one aspect, hybridization with probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins closely related molecules, may be used to identify nucleic acid sequences which encode casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins. The specificity of the probe, whether it is made from a highly specific region, e.g., unique nucleotides in the 5' regulatory region, or a less specific region, e.g., especially in the 3' coding region, and the stringency of the hybridization or amplification (maximal, high, intermediate, or low) will determine whether the probe identifies only naturally occurring sequences encoding casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins, alleles, or related sequences. Probes may also be used for the detection of related sequences, and should preferably contain at least 50% of the nucleotides from any of the casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and derived from the nucleotide sequence of the polynucleotide comprising (i) Drosophila gilgamesh (gish), human casein kinase 1, gamma 1, human casein kinase

1, gamma 2, human casein kinase 1, gamma 3 , (ii) Drosophila Gadfly Accession Number CG1534, human GABARAP, human GABARAP like 1, human GABARAP like, human GABARAP like 3, (iii) Drosophila Gadfly Accession Number CG10576, human PA2G4, (iv) Drosophila Mocs1, human MOCSA, human MOCS1 isoform 1, human MOCS1 isoform 2, human MOCS1 isoform 3, (v) Drosophila peanut (pnut), human CDC10, (vi) Drosophila Gadfly Accession Number CG7069, human pyruvate kinase, muscle, human pyruvate kinase, liver and RBC, (vii) Drosophila calreticulin (Crc), human calreticulin, or human calreticulin 2, or from genomic sequence including promoter, enhancer elements, and introns of the naturally occurring casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins. Means for producing specific hybridization probes for DNAs encoding casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins include the cloning of nucleic acid sequences encoding casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous protein derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, radionuclides such as ^{32}P or ^{35}S, or enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins may be used for the diagnosis of conditions or diseases, which are associated with expression of casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins. Examples of such conditions or diseases include, but are not limited to, pancreatic diseases and disorders, including diabetes. Polynucleotide sequences encoding

casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins may also be used to monitor the progress of patients receiving treatment for pancreatic diseases and disorders, including diabetes. The polynucleotide sequences encoding casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins may be used in Southern or Northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; or in dip stick, pin, ELISA or chip assays utilizing fluids or tissues from patient biopsies to detect altered casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous protein expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins may be useful in assays that detect activation or induction of various metabolic diseases such as obesity as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, and gallstones. The nucleotide sequences encoding casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins may be labeled by standard methods, and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantitated and compared with a standard value. The presence of altered levels of nucleotide sequences encoding casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins in the sample indicates the presence of the associated disease. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or in monitoring the treatment of an individual patient.

In order to provide a basis for the diagnosis of disease associated with expression of casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, which encodes casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with those from an experiment where a known amount of a substantially purified polynucleotide is used. Standard values obtained from normal samples may be compared with values obtained from samples from patients who are symptomatic for disease. Deviation between standard and subject values is used to establish the presence of disease. Once disease is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to evaluate whether the level of expression in the patient begins to approximate that, which is observed in the normal patient. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to metabolic diseases such as described above, the presence of a relatively high amount of transcript in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the pancreatic diseases and disorders.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding casein kinase 1 gamma, GABARAP, PA2G4, MOCS1,

CDC10, PK, calreticulin, or homologous proteins may involve the use of PCR. Such oligomers may be chemically synthesized, generated enzymatically, or produced from a recombinant source. Oligomers will preferably consist of two nucleotide sequences, one with sense orientation (5'.fwdarw.3') and another with antisense (3'.rarw.5'), employed under optimized conditions for identification of a specific gene or condition. The same two oligomers, nested sets of oligomers, or even a degenerate pool of oligomers may be employed under less stringent conditions for detection and/or quantification of closely related DNA or RNA sequences.

Methods which may also be used to quantitate the expression of casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and standard curves onto which the experimental results are interpolated (Melby, P. C. et al. (1993) J. Immunol. Methods, 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236). The speed of quantification of multiple samples may be accelerated by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantification.

In another embodiment of the invention, the nucleic acid sequences, which encode casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, calreticulin, or homologous proteins, may also be used to generate hybridization probes, which are useful for mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome or to a specific region of the chromosome using well known techniques. Such techniques include FISH, FACS, or artificial chromosome constructions, such as yeast artificial chromosomes, bacterial artificial chromosomes, bacterial P1 constructions or single chromosome cDNA libraries as reviewed in Price, C. M. (1993) Blood Rev. 7:127-134, and Trask, B. J. (1991) Trends Genet. 7:149-154. FISH (as described in Verma

et al. (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York, N.Y.) may be correlated with other physical chromosome mapping techniques and genetic map data. Examples of genetic map data can be found in the 1994 Genome Issue of Science (265:1981f). Correlation between the location of the gene encoding casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin on a physical chromosomal map and a specific disease, or predisposition to a specific disease, may help to delimit the region of DNA associated with that genetic disease.

The nucleotide sequences of the subject invention may be used to detect differences in gene sequences between normal, carrier, or affected individuals. In situ hybridisation of chromosomal preparations and physical mapping techniques such as linkage analysis using established chromosomal markers may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms, or parts thereof, by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localised by genetic linkage to a particular genomic region, for example, AT to 11q22-23 (Gatti, R. A. et al. (1988) Nature 336:577-580), any sequences mapping to that area may represent associated or regulatory genes for further investigation. The nucleotide sequences of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc. among normal, carrier, or affected individuals.

In another embodiment of the invention, the proteins, their catalytic or immunogenic fragments or oligopeptides thereof, an in vitro model, a genetically altered cell or animal, can be used for screening libraries of

compounds in any of a variety of drug screening techniques. One can identify effectors, e.g. receptors, enzymes, proteins, ligands, or substrates that bind to, modulate or mimic the action of one or more of the proteins of the invention. The protein or fragment thereof employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes, between the protein and the agent tested, may be measured. Agents could also, either directly or indirectly, influence the activity of the proteins of the invention.

Candidate agents may also be found in kinase assays where a kinase substrate such as a protein or a peptide, which may or may not include modifications as further described below, or others are phosphorylated by the proteins or protein fragments of the invention. A therapeutic candidate agent may be identified by its ability to increase or decrease the enzymatic activity of the proteins of the invention. The kinase activity may be detected by change of the chemical, physical or immunological properties of the substrate due to phosphorylation.

One example could be the transfer of radioisotopically labelled phosphate groups from an appropriate donor molecule to the kinase substrate catalyzed by the polypeptides of the invention. The phosphorylation of the substrate may be followed by detection of the substrates autoradiography with techniques well known in the art.

Yet in another example, the change of mass of the substrate due to its phosphorylation may be detected by mass spectrometry techniques.

One could also detect the phosphorylation status of a substrate with an analyte discriminating between the phosphorylated and unphosphorylated status of the substrate. Such an analyte may act by having different affinities for the phosphorylated and unphosphorylated forms of the

substrate or by having specific affinity for phosphate groups. Such an analyte could be, but is not limited to an antibody or antibody derivative, a recombinant antibody-like structure, a protein, a nucleic acid, a molecule containing a complexed metal ion, an anion exchange chromatography matrix, an affinity chromatography matrix or any other molecule with phosphorylation dependend selectivity towards the substrate.

Such an analyte could be employed to detect the kinase substrate, which is immobilized on a solid support during or after an enzymatic reaction. If the analyte is an antibody, its binding to the substrate could be detected by a variety of techniques as they are described in Harlow and Lane, 1998, Antibodies, CSH Lab Press, NY. If the analyte molecule is not an antibody, it may be detected by virtue of its chemical, physical or immunological properties, being endogenously associated with it or engineered to it.

Yet in another example the kinase substrate may have features, designed or endogenous, to facilitate its binding or detection in order to generate a signal that is suitable for the analysis of the substrates phosphorylation status. These features may be, but are not limited to a biotin molecule or derivative thereof, a glutathione-S-transferase moiety, a moiety of six or more consecutive histidine residues, an amino acid sequence or hapten to function as an epitope tag, a fluorochrome, an enzyme or enzyme fragment. The kinase substrate may be linked to these or other features with a molecular spacer arm to avoid steric hindrance.

In one example the kinase substrate may be labelled with a fluorochrome. The binding of the analyte to the labelled substrate in solution may be followed by the technique of fluorescence polarization as it is described in the literature (see, for example, Deshpande, S. et al. (1999) Prog. Biomed. Optics (SPIE) 3603:261; Parker, G. J. et al. (2000) J. Biomol. Screen. 5:77-88; Wu, P. et al. (1997) Anal. Biochem. 249:29-36).In a variation of this example, a fluorescent tracer molecule may compete with the

substrate for the analyte to detect kinase activity by a technique which is know to those skilled in the art as indirect fluorescence polarization.

In vivo, the enzymatic kinase activity of the unmodified polypeptides of casein kinase 1 gamma and pyruvate kinase towards a substrate can be enhanced by appropriate stimuli, triggering the phosphorylation of casein kinase 1 gamma and pyruvate kinase. This may be induced in the natural context by extracellular or intracellular stimuli, such as signaling molecules or environmental influences. One may generate a system containing activated casein kinase 1 gamma and pyruvate kinase, may it be an organism, a tissue, a culture of cells or cell-free environment, by exogenously applying this stimulus or by mimicking this stimulus by a variety of the techniques, some of them described further below. A system containing activated casein kinase 1 gamma and pyruvate kinase may be produced (i) for the purpose of diagnosis, study, prevention, and treatment of diseases and disorders related to body-weight regulation and thermogenesis, for example, but not limited to, metabolic diseases such as obesity, as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, and gallstones.

In addition activity of casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin against its physiological substrate(s) or derivatives thereof could be measured in cell-based assays. Agents may also interfere with posttranslational modifications of the protein, such as phosphorylation and dephosphorylation, farnesylation, palmitoylation, acetylation, alkylation, ubiquitination, proteolytic processing, subcellular localization and degradation. Moreover, agents could influence the dimerization or oligomerization of the proteins of the invention or, in a heterologous manner, of the proteins of the invention with other proteins, for example, but not exclusively, docking proteins, enzymes, receptors, or translation factors. Agents could also act on the physical interaction of the

proteins of this invention with other proteins, which are required for protein function, for example, but not exclusively, their downstream signaling.

Methods for determining protein-protein interaction are well known in the art. For example binding of a fluorescently labeled peptide derived from the interacting protein to the protein of the invention, or vice versa, could be detected by a change in polarisation. In case that both binding partners, which can be either the full length proteins as well as one binding partner as the full length protein and the other just represented as a peptide are fluorescently labeled, binding could be detected by fluorescence energy transfer (FRET) from one fluorophore to the other. In addition, a variety of commercially available assay principles suitable for detection of protein-protein interaction are well known In the art, for example but not exclusively AlphaScreen (PerkinElmer) or Scintillation Proximity Assays (SPA) by Amersham. Alternatively, the interaction of the proteins of the invention with cellular proteins could be the basis for a cell-based screening assay, in which both proteins are fluorescently labeled and interaction of both proteins is detected by analysing cotranslocation of both proteins with a cellular imaging reader, as has been developed for example, but not exclusively, by Cellomics or EvotecOAI. In all cases the two or more binding partners can be different proteins with one being the protein of the invention, or in case of dimerization and/or oligomerization the protein of the invention itself. Proteins of the invention, for which one target mechanism of interest, but not the only one, would be such protein/protein interactions are casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, *CDC10, PK, and calreticulin.*

Assays for determining enzymatic activity of the proteins of the invention are well known in the art.

Of particular interest are screening assays for agents that have a low toxicity for mammalian cells. The term "agent" as used herein describes

any molecule, e.g. protein or pharmaceutical, with the capability of altering or mimicking the physiological function of one or more of the proteins of the invention. Candidate agents encompass numerous chemical classes, though typically they are organic molecules, preferably small organic compounds having a molecular weight of more than 50 and less than about 2,500 Daltons. Candidate agents comprise functional groups necessary for structural interaction with proteins, particularly hydrogen bonding, and typically include at least an amine, carbonyl, hydroxyl or carboxyl group, preferably at least two of the functional chemical groups. The candidate agents often comprise carbocyclic or heterocyclic structures and/or aromatic or polyaromatic structures substituted with one or more of the above functional groups.

Candidate agents are also found among biomolecules including peptides, saccharides, fatty acids, steroids, purines, pyrimidines, nucleic acids and derivatives, structural analogs or combinations thereof. Candidate agents are obtained from a wide variety of sources including libraries of synthetic or natural compounds. For example, numerous means are available for random and directed synthesis of a wide variety of organic compounds and biomolecules, including expression of randomized oligonucleotides and oligopeptides. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts are available or readily produced. Additionally, natural or synthetically produced libraries and compounds are readily modified through conventional chemical, physical and biochemical means, and may be used to produce combinatorial libraries. Known pharmacological agents may be subjected to directed or random chemical modifications, such as acylation, alkylation, esterification, amidification, etc. to produce structural analogs. Where the screening assay is a binding assay, one or more of the molecules may be joined to a label, where the label can directly or indirectly provide a detectable signal.

Another technique for drug screening, which may be used, provides for high throughput screening of compounds having suitable binding affinity to the protein of interest as described in published PCT application WO84/03564. In this method, as applied to the protein of the invention large numbers of different small test compounds, e.g. aptamers, peptides, low-molecular weight compounds, etc. are synthesized on a solid substrate, such as plastic pins or some other surface. The test compounds are reacted with the protein, or fragments thereof, and washed. Bound proteins are then detected by methods well known in the art. Purified proteins can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilise it on a solid support. In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding the protein specifically compete with a test compound for binding the protein. In this manner, the antibodies can be used to detect the presence of any peptide, which shares one or more antigenic determinants with the protein of the invention.

The nucleic acids encoding the proteins of the invention can be used to generate transgenic cell lines and animals. These transgenic non-human animals are useful in the study of the function and regulation of the proteins of the invention in vivo. Transgenic animals, particularly mammalian transgenic animals, can serve as a model system for the investigation of many developmental and cellular processes common to humans. A variety of non-human models of metabolic disorders can be used to test modulators of the protein of the invention. Misexpression (for example, overexpression or lack of expression) of the protein of the invention, particular feeding conditions, and/or administration of biologically active compounts can create models of metablic disorders.

In one embodiment of the invention, such assays use mouse models of insulin resistance and/or diabetes, such as mice carrying gene knockouts in the leptin pathway (for example, ob (leptin) or db (leptin receptor) mice). Such mice develop typical symptoms of diabetes , show hepatic lipid accumulation and frequently have increased plasma lipid levels (see Bruning et al, 1998, Mol. Cell. 2:449-569). Susceptible wild type mice (for example C57Bl/6) show similiar symptoms if fed a high fat diet. In addition to testing the expression of the proteins of the invention in such mouse strains (see EXAMPLE 4), these mice could be used to test whether administration of a candidate modulator alters for example lipid accumulation in the liver, in plasma, or adipose tissues using standard assays well known in the art, such as FPLC, colorimetric assays, blood glucose level tests, insulin tolerance tests and others.

Transgenic animals may be made through homologous recombination in embryonic stem cells, where the normal locus of the gene encoding the protein of the invention is mutated. Alternatively, a nucleic acid construct encoding the protein is injected into oocytes and is randomly integrated into the genome. One may also express the genes of the invention or variants thereof in tissues where they are not normally expressed or at abnormal times of development. Furthermore, variants of the genes of the invention like specific constructs expressing anti-sense molecules or expression of dominant negative mutations, which will block or alter the expression of the proteins of the invention may be randomly integrated into the genome. A detectable marker, such as lac Z or luciferase may be introduced into the locus of the genes of the invention, where upregulation of expression of the genes of the invention will result in an easily detectable change in phenotype. Vectors for stable integration include plasmids, retroviruses and other animal viruses, yeast artificial chromosomes (YACs), and the like. DNA constructs for homologous recombination will contain at least portions of the genes of the invention with the desired genetic modification, and will include regions of homology

to the target locus. Conveniently, markers for positive and negative selection are included. DNA constructs for random integration do not need to contain regions of homology to mediate recombination. DNA constructs for random integration will consist of the nucleic acids encoding the proteins of the invention, a regulatory element (promoter), an intron and a poly-adenylation signal. Methods for generating cells having targeted gene modifications through homologous recombination are known in the field. For embryonic stem (ES) cells, an ES cell line may be employed, or embryonic cells may be obtained freshly from a host, e.g. mouse, rat, guinea pig, etc. Such cells are grown on an appropriate fibroblast-feeder layer and are grown in the presence of leukemia inhibiting factor (LIF). ES or embryonic cells may be transfected and can then be used to produce transgenic animals. After transfection, the ES cells are plated onto a feeder layer in an appropriate medium. Cells containing the construct may be selected by employing a selection medium. After sufficient time for colonies to grow, they are picked and analyzed for the occurrence of homologous recombination. Colonies that are positive may then be used for embryo manipulation and morula aggregation. Briefly, morulae are obtained from 4 to 6 week old superovulated females, the Zona Pellucida is removed and the morulae are put into small depressions of a tissue culture dish. The ES cells are trypsinized, and the modified cells are placed into the depression closely to the morulae. On the following day the aggregates are transfered into the uterine horns of pseudopregnant females. Females are then allowed to go to term. Chimeric offsprings can be readily detected by a change in coat color and are subsequently screened for the transmission of the mutation into the next generation (F1-generation). Offspring of the F1-generation are screened for the presence of the modified gene and males and females having the modification are mated to produce homozygous progeny. If the gene alterations cause lethality at some point in development, tissues or organs can be maintained as allogenic or congenic grafts or transplants, or in vitro culture. The transgenic animals may be any non-human mammal, such as laboratory animal, domestic

animals, etc., for example, mouse, rat, guinea pig, sheep, cow, pig, and others. The transgenic animals may be used in functional studies, drug screening, and other applications and are useful in the study of the function and regulation of the proteins of the invention in vivo.

Finally, the invention also relates to a kit comprising at least one of

(a) a casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin nucleic acid molecule or a fragment thereof;

(b) a casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin amino acid molecule or a fragment or an isoform thereof;

(c) a vector comprising the nucleic acid of (a);

(d) a host cell comprising the nucleic acid of (a) or the vector of (b);

(e) a polypeptide encoded by the nucleic acid of (a);

(f) a fusion polypeptide encoded by the nucleic acid of (a);

(g) an antibody, an aptamer or another receptor against the nucleic acid of (a) or the polypeptide of (b), (e) or (f) and

(h) an anti-sense oligonucleotide of the nucleic acid of (a).

The kit may be used for diagnostic or therapeutic purposes or for screening applications as described above. The kit may further contain user instructions.

The Figures show:

Figure 1 shows the triglyceride content of a gilgamesh casein kinase 1 (gish; Gadfly Accession Number CG6963) mutant. Shown is the increase of triglyceride content of HD-EP(3)37409 flies (referred to as "HD-EP37409" in column 2) caused by homozygous viable integration of the P-vector into the promotor region of the second transcription unit of gilgamesh, in comparison to controls with integration of this vector type

(referred to as "EP-control" in column 1). Also shown is that ectopic expression of the gilgamesh gene mainly in the fat body of the flies (referred to as "HD-EP37409/FB" in column 4) in comparison to controls with integration of this vector type (referred to as "random EP/FB" in column 3) causes no change of triglyceride content, and that ectopic expression of the gilgamesh gene mainly in the neurons of the flies (referred to as "HD-EP37409/elav" in column 6) in comparison to controls with integration of this vector type (referred to as "random EP/elav" in column 5) causes a decrease of triglyceride content.

Figure 2 shows the molecular organisation of the mutated gilgamesh casein kinase 1 (Gadfly Accession Number CG6963) gene locus.

Figure 3 shows the human homologs of Gadfly Accession Number CG6963 (gilgamesh)

Figure 3A. BLASTP search result for Gadfly Accession Number CG6963 (Query) with the best human homolog match (Sbject)

Figure 3B shows the nucleotide sequence encoding human casein kinase 1, gamma 1 (Genbank Accession Number AB042563; SEQ ID NO:1)

Figure 3C shows the amino acid sequence of human casein kinase 1, gamma 1 (Genbank Accession Number Q9HCP0; SEQ ID NO:2)

Figure 3D shows the nucleotide sequence encoding human casein kinase 1, gamma 2 (Genbank Accession Number NM_001319; SEQ ID NO:3)

Figure 3E shows the amino acid sequence of human casein kinase 1, gamma 2 (Genbank Accession Number NP_001310; SEQ ID NO:4)

Figure 3F shows the nucleotide sequence encoding human casein kinase 1, gamma 3 (Genbank Accession Number NM_004384; SEQ ID NO:5)

Figure 3G shows the amino acid sequence of human casein kinase 1, gamma 3 (Genbank Accession Number NP_004375; SEQ ID NO:6).

Figure 4 shows the comparison (Clustal W (1.83) protein sequence alignment analysis) of human and Drosophila casein kinase 1 proteins.

Gaps in the alignment are represented as -. In the figure 'CK1 g3 Hs' refers to human casein kinase 1 gamma 3, 'CK1 g1 Hs' refers to human casein kinase 1 gamma 1, 'CK1 g2 Hs' refers to human casein kinase 1 gamma 2, and 'CG6963 Dm' refers to the Drosophila gilgamesh gene product with Gadfly Accession Number CG6963.

Figure 5 shows the analysis of casein kinase 1, gamma 1 and casein kinase 1, gamma 3 expression in mammalian tissues. The relative RNA-expression is shown on the Y-axis, the tissues tested are given on the X-axis. WAT refers to white adipose tissue, BAT refers to brown adipose tissue.

Figure 5A shows the real-time PCR analysis of casein kinase 1, gamma 1 expression in mouse wildtype tissues.

Figure 5B shows the real-time PCR analysis of casein kinase 1, gamma 1 expression in wildtype mice (WT-mice), compared to genetically obese mice (ob/ob-mice) and to fasted mice (fasted-mice).

Figure 5C shows the real-time PCR analysis of casein kinase 1, gamma 3 expression in mouse wildtype tissues. WAT refers to white adipose tissue, BAT refers to brown adipose tissue.

Figure 5D shows the real-time PCR analysis of casein kinase 1, gamma 3 expression in wildtype mice (WT-mice), compared to genetically obese mice (ob/ob-mice) and to fasted mice (fasted-mice).

Figure 6 shows the decrease of triglyceride content of PX6298.1 flies caused by integration of the P-vector (in comparison to controls with integration of these vectors elsewhere in genome).

Figure 7 shows the molecular organisation of the mutated GABARAP (Gadfly Accession Number CG1534) gene locus. The Gadfly Accession Number CG1534 annotated gene encodes three different transcripts. Only one of these transcripts encodes GABARAP (Gadfly Accession Number

CT3947; synonym for Gadfly Accession Number CG32672 and Genbank Accession Number. NM_167245).

Figure 8 shows the human homologs of Gadfly Accession Number CG1534
Figure 8A. BLASTP search result for Gadfly Accession Number CG1534 (Query) with the best human homolog match (Sbject)
Figure 8B shows the nucleotide sequence encoding human GABARAP (Genbank Accession Number NM_007278; SEQ ID NO:7)
Figure 8C shows the amino acid sequence of human GABARAP (Genbank Accession Number NP_009209; SEQ ID NO:8)
Figure 8D shows the nucleotide sequence encoding GABARAP like 1 (Genbank Accession Number NM_031412; SEQ ID NO:9)
Figure 8E shows the amino acid sequence of human GABARAP like 1 (Genbank Accession Number NP_113600; SEQ ID NO:10)
Figure 8F shows the nucleotide sequence encoding human GABARAP like 2 (Genbank Accession Number NM_007285; SEQ ID NO:11)
Figure 8G shows the amino acid sequence of human GABARAP like 2 (Genbank Accession Number NP_009216; SEQ ID NO:12)
Figure 8H shows the nucleotide sequence encoding human GABARAP like 3 (Genbank Accession Number NM_032568; SEQ ID NO:13)
Figure 8I shows the amino acid sequence of human GABARAP like 3 (Genbank Accession Number NP_115957; SEQ ID NO:14)

Figure 9 shows a comparison (Clustal W (1.82) protein sequence alignment analysis) of human and Drosophila GABARAP proteins. Gaps in the alignment are represented as -. In the figure 'GABARAP-I3 Hs' refers to human GABARAP like 3, 'GABARAP-I1 Hs' refers to human GABARAP like 1, 'CG1534 Dm' refers to Drosophila protein encoded by Gadfly Accession Number CG1534, 'CG12334 Dm' refers to Drosophila protein encoded by Gadfly Accession Number CG12334, and 'GABARAP-I2 Hs' refers to the human GABARAP like 2.

Figure 10 shows the analysis of GABARAP 2 expression in mammalian tissues. The relative RNA-expression is shown on the Y-axis. In Figure 10A and 10B the tissues tested are given on the X-axis. WAT refers to white adipose tissue, BAT refers to brown adipose tissue. In Figure 10C, the X-axis represents the time axis. 'd0' refers to day 0 (start of the experiment), 'd2' - 'd10' refers to day 2 - day 10 of adipocyte differentiation).

Figure 10A shows the real-time PCR analysis of GABARAP 2 expression in mouse wildtype tissues.

Figure 10B shows the real-time PCR analysis of GABARAP 2 expression in wildtype mice (WT-mice), compared to genetically obese mice (ob/ob-mice) and to fasted mice (fasted-mice).

Figure 10C shows the real-time PCR analysis GABARAP 2 expression in mammalian fibroblast (3T3-F442A) cells, during the differentiation from preadipocytes to mature adipocytes.

Figure 11 shows the increase of triglyceride content of EP(3)3271 flies caused by homozygous viable integration of the P-vector into the first exon of Gadfly Accession Number CG10576 (in comparison to controls with integration of these vectors).

Figure 12 shows the molecular organisation of the mutated methionyl aminopeptidase (Gadfly Accession Number CG10576) gene locus.

Figure 13 shows the human homologs of Gadfly Accession Number CG10576

Figure 13A. shows the BLASTP search result for Gadfly Accession Number CG10576 (Query) with the best human homolog match (Sbject)

Figure 13B shows the nucleotide sequence encoding human proliferation associated protein 2G4, 38 kDa (Genbank Accession Number NM_006191; SEQ ID NO:15)

Figure 13C shows the amino acid sequence of human proliferation associated protein 2G4, 38 kDa (Genbank Accession Number NP_006182; SEQ ID NO:16).

Figure 14 shows the ClustalW (1.7) protein alignment for the Drosophila protein encoded by GadFly Accession Number CG10576 and human p38-2G4 (referred to as ' XP_049048.1').

Figure 15 shows the analysis of proliferation associated 2G4 protein, 38 kDa (PA2G4) expression in mammalian tissues. The relative RNA-expression is shown on the Y-axis, the tissues tested are given on the X-axis. WAT refers to white adipose tissue, BAT refers to brown adipose tissue.

Figure 15A shows the real-time PCR analysis of PA2G4 expression in mouse wildtype tissues.

Figure 15B shows the real-time PCR analysis of PA2G4 expression in wildtype mice (WT-mice), compared to genetically obese mice (ob/ob-mice) and to fasted mice (fasted-mice).

Figure 16 shows the increase of triglyceride content of EP(3)3688 flies caused by homozygous viable integration of the P-vector (in comparison to controls without integration of this vector).

Figure 17 shows the molecular organisation of the mutated Mocs1 (Gadfly Accession Number CG7858) gene locus.

Figure 18 shows the human homologs of Gadfly Accession Number CG7858 (Mocs1)

Figure 18A shows the BLASTP search results for Gadfly Accession Number CG7858 (Mocs1) (referred to as 'Query'), shown are only the human homologs (referred to as 'Sbjct') with highest homology values.

Figure 18B shows the nucleotide sequence encoding human MOCSA and MOCSC (Genbank Accession Number AF034374; SEQ ID NO:17)

Figure 18C shows the amino acid sequence of human MOCSA (Genbank Accession Number AAB87523; SEQ ID NO:18)

Figure 18D shows the nucleotide sequence encoding human MOCS1 protein, isoform 1 (Genbank Accession Number XM_166358; SEQ ID NO:19)

Figure 18E shows the amino acid sequence of human MOCS1 protein, isoform 1 (Genbank Accession Number XP_166358; SEQ ID NO:20)

Figure 18F shows the nucleotide sequence encoding human MOCS1, isoform 2 (Genbank Accession Number NM_005942; SEQ ID NO:21)

Figure 18G shows the amino acid sequence of human MOCS1, isoform 2 (Genbank Accession Number NP_005933; SEQ ID NO:22)

Figure 18H shows the nucleotide sequence encoding human MOCS1, isoform 3 (Genbank Accession Number NM_138928; SEQ ID NO:23)

Figure 18I shows the amino acid sequence of human MOCS1, isoform 3 (Genbank Accession Number NP_620306; SEQ ID NO:24).

Figure 19 shows the comparison (Clustal W (1.82) protein sequence alignment analysis) of human and Drosophila Mocs1 proteins. Gaps in the alignment are represented as -. In the figure 'Mocs1-2 Hs' refers to human Mocs1, isoform 2, 'Mocs1-3 Hs' refers to human Mocs1, isoform 3, 'Mocs1-1 Hs' refers to human Mocs1, isoform 1, 'Mocs1 Hs' refers to human MocsA, 'Mocs1-PA Dm' refers to Drosophila Mocs1 protein variant A, and 'Mocs1-PC Dm' refers to Drosophila Mocs1 protein variant C.

Figure 20 shows the analysis of Mocs expression in mammalian tissues. The relative RNA-expression is shown on the Y-axis, the tissues tested are given on the X-axis. WAT refers to white adipose tissue, BAT refers to brown adipose tissue.

Figure 20A shows the real-time PCR analysis of Mocs expression in mouse wildtype tissues.

Figure 20B shows the real-time PCR analysis of Mocs expression in wildtype mice (WT-mice), compared to genetically obese mice (ob/ob-mice) and to fasted mice (fasted-mice).

Figure 21 shows the triglyceride content of a peanut protein (pnut; Gadfly Accession Number CG8705) mutant. Shown is the increase of triglyceride content of EP(2)2036 flies caused by homozygous viable integration of the P-vector into the promoter/enhancer of peanut (in comparison to controls -EP control- with integration of these vectors).

Figure 22 shows the molecular organisation of the mutated pnut (Gadfly Accession Number CG8705) gene locus.

Figure 23 shows the human homologs of Gadfly Accession Number CG8705 (peanut)

Figure 23A shows the BLASTP search results for Gadfly Accession Number CG8705

Figure 23B shows the nucleotide sequence encoding human CDC10 cell division cycle 10 homolog (Genbank Accession Number NM_001788; SEQ ID NO:25)

Figure 23C shows the amino acid sequence of human CDC10 cell division cycle 130 homolog (Genbank Accession Number NP_001779; SEQ ID NO:26).

Figure 24 shows the ClustalW (1.7) protein sequence alignment for Gadfly Accession Number CG8705, human CDC10 ('XM_011595'), and human CDC10 homolog (septin) ('NM_001788')

Figure 25 shows the analysis of the peanut homolog (referred to as 'Peanut') expression in mammalian tissues. The relative RNA-expression is shown on the Y-axis, the tissues tested are given on the X-axis. WAT refers to white adipose tissue, BAT refers to brown adipose tissue.

Figure 25A shows the real-time PCR analysis of Peanut expression in mouse wildtype tissues.

Figure 25B shows the real-time PCR analysis of Peanut expression in wildtype mice (WT-mice), compared to genetically obese mice (ob/ob-mice) and to fasted mice (fasted-mice).

Figure 26 shows the triglyceride content of a pyruvate kinase protein (Gadfly Accession Number CG7069) mutant. Shown is the increase of triglyceride content of EP(3)3224 flies caused by homozygous viable integration of the P-vector into the second exon of the pyruvate kinase gene (in comparison to controls -EP control- with integration of these vectors).

Figure 27 shows the molecular organization of the mutated pyruvate kinase (Gadfly Accession Number CG7069) gene locus.

Figure 28 shows the human homologs of Gadfly Accession Number CG7069

Figure 28A shows the BLASTP search result for Gadfly Accession Number CG7069 (Query) with the best human homologous match (Sbject).

Figure 28B shows the nucleotide sequence encoding human pyruvate kinase, muscle (Genbank Accession Number X56494; SEQ ID NO:27)

Figure 28C shows the amino acid sequence of human pyruvate kinase, muscle, M1 isozyme (Genbank Accession Number P14618; SEQ ID NO:28)

Figure 28D shows the amino acid sequence of human pyruvate kinase, muscle, M2 isozyme (Genbank Accession Number P14786; SEQ ID NO:29)

Figure 28E shows the nucleotide sequence encoding human pyruvate kinase, liver and RBC (Genbank Accession Number NM_000298; SEQ ID NO:30)

Figure 28F shows the amino acid sequence of human pyruvate kinase, liver and RBC (Genbank Accession Number NP_000289; SEQ ID NO:31).

Figure 29 shows the ClustalW (1.7) protein sequence alignment analysis of Drosophila, mouse, and human pyruvate kinase. In the figure 'pk3_h2' refers to human pyruvate kinase, muscle (Genbank Accession Number. NP_002645), 'pk3_h' refers to human pyruvate kinase, muscle (Genbank Accession Number XM_037768), 'pk3_m' refers to mouse pyruvate kinase 3 (Genbank Accession Number BC016619), and 'pk3_dro' refers to Drosophila pyruvate kinase (Gadfly Accession Number CG7069).

Figure 30 shows the increase of triglyceride content of EP(3)3321, EP(3)0834, and EP(3)0979 flies caused by homozygous viable integration of the P-vector in the transcription unit of Gadfly Accession Number CG9429 (in comparison to controls without integration of this vector).

Figure 31 shows the molecular organisation of the calreticulin (Crc; Gadfly Accession Number CG9429) gene locus.

Figure 32 shows the human homologs of Gadfly Accession Number CG9429 (calreticulin)
Figure 32A shows the BLASTP search result for Gadfly Accession Number CG9429 (Query) with the best human homologous match (Sbject).
Figure 32B shows the nucleotide sequence encoding human Calreticulin (Genbank Accession Number NM_004343; SEQ ID NO:32)
Figure 32C shows the amino acid sequence of human Calreticulin (Genbank Accession Number NP_004334; SEQ ID NO:33)
Figure 32D shows the nucleotide sequence encoding human Calreticulin 2 (Genbank Accession Number NM_145046; SEQ ID NO:34)
Figure 32E shows the amino acid sequence of human Calreticulin 2 (Genbank Accession Number NP_659483; SEQ ID NO:35).

Figure 33 shows the comparison (Clustal W (1.82) protein sequence alignment analysis) of human and Drosophila calreticulin proteins. Gaps in the alignment are represented as -. In the figure 'crc Dm' refers to

Drosophila calreticulin, 'crc Hs' refers to human calreticulin, and 'MGC26577 Hs' refers to human calreticulin 2.

The examples illustrate the invention:

Example 1: Measurement of triglyceride content

Mutant flies are obtained from a fly mutation stock collection. The flies are grown under standard conditions known to those skilled in the art. In the course of the experiment, additional feedings with bakers yeast (Saccharomyces cerevisiae) are provided. The average increase of triglyceride content of Drosophila containing the EP-vectors in homozygous or hemizygous viable integration was investigated in comparison to control flies (see FIGURES 1, 6, 11, 16, 21, 26, and 30). For determination of triglyceride, flies were incubated for 5 min at preferably 90°C in an aqueous buffer using a waterbath, followed by hot extraction. After another 5 min incubation at preferably 90°C and mild centrifugation, the triglyceride content of the flies extract was determined using Sigma Triglyceride (INT 336-10 or -20) assay by measuring changes in the optical density according to the manufacturer's protocol. As a reference protein content of the same extract was measured using BIO-RAD DC Protein Assay according to the manufacturer's protocol. The assay was repeated several times.

The average triglyceride level of all flies of the EP collections (referred to as 'EP-control') is shown as 100% (ratio triglyceride content/protein content) in the first columns in FIGURES 1, 11, 16, 21, 26, and 30, including standard deviation. The average triglyceride level of all flies of the PX collection (referred to as 'PX-lines') is shown as 1 (relative amount of triglyceride/fly) in the first column in FIGURE 6, including standard deviation. The average triglyceride level of all flies containing the FB- Gal4

vector (referred to as 'random EP/FB') is shown as 100% (ratio triglyceride content/protein content) in the third column in FIGURE 1. The average triglyceride level of all flies containing the elav- Gal4 vector (referred to as 'random EP/elav') is shown as 100% in the fifth column in FIGURE 1.

HD-EP(3)37409 homozygous flies show constantly a higher triglyceride content than the controls (142 %; column 2 in FIGURE 1). The offspring of HD-EP(3)37409 males that are crossed to FB-Gal4 virgins, carrying a copy of the HD-EP(3)37409 vector and a copy of the FB-Gal4 vector, leading to ectopic expression of adjacent genomic DNA sequences 3' of the HD-EP(3)37409 integration locus, mainly in the fatbody of these flies, show no changes in triglyceride content compared with the controls (103 %, column 4 in Figure 1). The offspring of HD-EP(3)37409 males that are crossed to elav-Gal4 virgins, carrying a copy of the HD-EP(3)37409 vector and a copy of the elav-Gal4 vector, leading to ectopic expression of adjacent genomic DNA sequences 3' of the HD-EP(3)37409 integration locus, mainly in the neurons of these flies, show constantly a lower triglyceride content than the controls (70%; colunm 6 in Figure 1). Therefore, the loss of the gene activity and the gain of gene activity in the locus 98B17-19 on chromosome 3R where the EP-vector of HD-EP(3)37409 flies is homozygous viable integrated 5' of the gilgamesh gene, are in both cases responsible for changes in the metabolism of the energy storage triglycerides.

PX6298.1 hemizygous flies show constantly a lower triglyceride content than the controls (column 2 in FIGURE 6). Therefore, the change of gene activity in the locus of the PX6298.1 integration on chromosome X where the PX-vector of PX6298.1 flies is hemizygous viable integrated, is responsible for changes in the metabolism of the energy storage triglycerides.

EP(3)3271 homozygous flies show constantly a higher triglyceride content than the controls (column 2 in FIGURE 11). Therefore, the loss of gene activity in the locus 64F1 on chromosome 3L where the EP-vector of HD-EP(3)3271 flies is homozygous viable integrated, is responsible for changes in the metabolism of the energy storage triglycerides.

EP(3)3688 homozygous flies show constantly a higher triglyceride content than the controls (column 2 in FIGURE 16). Therefore, the loss of gene activity in the locus 68A3-68A3 on chromosome 3L where the EP-vector of EP(3)3688 flies is homozygous viable integrated, is responsible for changes in the metabolism of the energy storage triglycerides.

EP(2)2036 homozygous flies show constantly a higher triglyceride content than the controls (column 2 in FIGURE 21). Therefore, the loss of gene activity in the locus 44B3-44B4 on chromosome 2L where the EP-vector of EP(2)2036 flies is homozygous viable integrated, is responsible for changes in the metabolism of the energy storage triglycerides.

EP(3)3224 homozygous flies show constantly a higher triglyceride content than the controls (153%; column 2 in FIGURE 26). Therefore, the loss of gene activity in the locus 94A15-16 on chromosome 3R where the EP-vector of EP(3)3224 flies is homozygous viable integrated, is responsible for changes in the metabolism of the energy storage triglycerides.

EP(3)3321, EP(3)0834, and EP(3)0979 homozygous flies show constantly a higher triglyceride content than the controls (columns 2 to 4 in FIGURE 30). Therefore, the loss of gene activity in the locus 85E2 on chromosome 3R where the EP-vectors of EP(3)3321, EP(3)0979, or EP(3)0834 flies are homozygous viable integrated, is responsible for changes in the metabolism of the energy storage triglycerides.

Example 2: Identification of Drosophila genes and proteins associated with metabolic control

Nucleic acids encoding the proteins of the present invention were identified using a plasmid-rescue technique. Genomic DNA sequences were isolated that are localized to the EP vector (herein HD-EP(3)37409, PX6298.1, EP(3)3271, EP(3)3688, EP(2)2036, EP(3)3224, EP(3)3321, EP(3)0834, and EP(3)0979) integration. Using those isolated genomic sequences public databases like Berkeley Drosophila Genome Project (GadFly) were screened thereby identifying the integration sites of the vectors, and the corresponding genes. The molecular organization of these gene loci is shown in FIGURES 2, 7, 12, 17, 22, 27, and 31.

In FIGURE 2, genomic DNA sequence is represented by the assembly as a thin black line in the middle (numbers represent the length in basepairs of the genomic DNA) that includes the integration sites of vector for line HD-EP(3)37409. Transcribed DNA sequences (ESTs) and predicted exons are shown as bars on the two sides (sense and antisense strand). Predicted exons of the cDNA with GadFly Accession Number CG6963 (referred to as gilgamesh or gish) are shown as dark grey bars and introns as light grey lines. The sequence encodes for a gene that is predicted by GadFly sequence analysis programs as Accession Number CG6963. Public DNA sequence databases (for example, NCBI GenBank) were screened thereby identifying the integration sites of lines HD-EP(3)37409, causing an increase of triglyceride content. HD-EP(3)37409 is integrated into the promoter region of the second transcription unit in sense orientation of the cDNA with GadFly Accession Number CG6963 (the site of integration is shown as vertical dotted line). Therefore, expression of the cDNA encoding gilgamesh could be effected by integration of vectors of line HD-EP(3)37409, or gilgamesh could be ectopically expressed, e.g. in neurons, leading to a change of the energy storage triglycerides.

In FIGURE 7, genomic DNA sequence is represented by the assembly as a thin black line in the middle (numbers represent the length in basepairs of the genomic DNA) that includes the integration sites of vector for line PX6298.1. Transcribed DNA sequences (ESTs) and predicted exons are shown as bars on the two sides (sense and antisense strand). Predicted exons of the cDNA with GadFly Accession Number CG1534 are shown as dark grey bars and introns as light grey lines. The sequence encodes for a gene that is predicted by GadFly sequence analysis programs as Accession Number CG1534. The integration site of the vector for line PX6298.1 was identified at position 215791 on Drosophila chromosome X. Predicted exons of the cDNA with GadFly Accession Number CG1534 are located on chromosome X in three positions, starting with ATG start codons at positions 209052 (two transcripts), and 215668. Only the transcript with the start codon at position 215668 encodes for GABARAP (Gadfly Accession Number CT3947; synonym for Gadfly Accession Number CG32672 and Genbank Accession Number. NM_167245). Using those isolated genomic sequences public databases like Berkeley Drosophila Genome Project (GadFly) were screened confirming the hemizygous viable integration site of the PX6298.1 vector in the 5prime untranslated region of the first exon of the gene encoding GABARAP (123 base pairs 5prime of the start codon), causing a decrease of triglyceride content.

In FIGURE 12, genomic DNA sequence is represented by the assembly as a dotted black line (from position 5703500 to 5707500 on chromosome 3L) that includes the integration sites of vector for line EP(3)3271. Transcribed DNA sequences (ESTs) and predicted exons are shown as bars in the lower two lines. Predicted exons of the cDNA with GadFly Accession Number CG10576 are shown as dark grey bars and introns as light grey bars. Methionyl aminopeptidase encodes for a gene that is predicted by GadFly sequence analysis programs as Accession Number CG10576. Public DNA sequence databases (for example, NCBI GenBank) were screened thereby identifying the integration sites of lines EP(3)3271,

causing an increase of triglyceride content. EP(3)3271 is integrated into the first exon in antisense orientation of the cDNA with Accession Number CG10576. Therefore, expression of the cDNA encoding Accession Number CG10576 could be effected by homozygous integration of vectors of line EP(3)3271, leading to increase of the energy storage triglycerides.

In FIGURE 17, genomic DNA sequence is represented by the assembly as a dotted black line (from position 10988000 to 10992000 on chromosome 3L) that includes the integration sites of vector for line EP(3)3688. Transcribed DNA sequences (ESTs) and predicted exons are shown as bars in the lower two lines. Predicted exons of the cDNA with GadFly Accession Number CG7858 are shown as dark grey bars and introns as light grey bars. Mocs1 encodes for a gene that is predicted by GadFly sequence analysis programs as Accession Number CG7858. Public DNA sequence databases (for example, NCBI GenBank) were screened thereby identifying the integration sites of lines EP(3)3688, causing an increase of triglyceride content. EP(3)3688 is integrated into the promoter in sense direction of the cDNA with Accession Number CG7858. Therefore, expression of the cDNA encoding Accession Number CG7858 could be effected by homozygous integration of vectors of line EP(3)3688, leading to increase of the energy storage triglycerides.

In FIGURE 22, genomic DNA sequence is represented by the assembly as a dotted black line (from position 3272156 to 3277156 on chromosome 2R) that includes the integration sites of vector for line EP(2)2036. Transcribed DNA sequences (ESTs) and predicted exons are shown as bars in the lower two lines. Predicted exons of the cDNA with GadFly Accession Number CG8705 are shown as dark grey bars and introns as light grey bars. Pnut encodes for a gene that is predicted by GadFly sequence analysis programs as Accession Number CG8705. Public DNA sequence databases (for example, NCBI GenBank) were screened thereby identifying the integration sites of lines EP(2)2036, causing an increase of

triglyceride content. EP(2)2036 is integrated into the promoter/enhancer of peanut in antisense orientation of the cDNA with Accession Number CG8705. Therefore, expression of the cDNA encoding Accession Number CG8705 could be effected by homozygous integration of vectors of line EP(2)2036, leading to increase of the energy storage triglycerides.

In FIGURE 27, genomic DNA sequence is represented by the assembly as a dotted black line (from position 18113034 to 18116159 on chromosome 3R) that includes the integration sites of vector for line EP(3)3224. Transcribed DNA sequences (ESTs) and predicted exons are shown as bars in the upper two lines. Predicted exons of the cDNA with GadFly Accession Number CG7069 are shown as dark grey bars and introns as light grey bars. The sequence encodes for a gene that is predicted by GadFly sequence analysis programs as Accession Number CG7069. Public DNA sequence databases (for example, NCBI GenBank) were screened thereby identifying the integration sites of lines EP(3)3224, causing an increase of triglyceride content. EP(3)3224 is integrated into the second exon of pyruvate kinase in sense orientation of the cDNA with GadFly Accession Number CG7069. Therefore, expression of the cDNA encoding GadFly Accession Number CG7069 could be effected by homozygous integration of vectors of line EP(3)3224, leading to an increase of the energy storage triglycerides.

In FIGURE 31, genomic DNA sequence is represented by the assembly as a dotted black line (from position 5435825 to 5438950 on chromosome 3R) that includes the integration sites of vectors for lines EP(3)3321, EP(3)0979, and EP(3)0834. Transcribed DNA sequences (ESTs) and predicted exons are shown as bars in the lower two lines. Predicted exons of the cDNA with GadFly Accession Number CG9429 are shown as dark grey bars and introns as light grey bars. calreticulin encodes for a gene that is predicted by GadFly sequence analysis programs as Accession Number CG9429. Public DNA sequence databases (for example, NCBI GenBank)

were screened thereby identifying the integration sites of lines EP(3)3321, EP(3)0979, and EP(3)0834, causing an increase of triglyceride content. EP(3)3321, EP(3)0979, and EP(3)0834 are integrated into the transcription unit of the cDNA with Accession Number CG9429. Therefore, expression of the cDNA encoding Accession Number CG9429 could be effected by homozygous integration of vectors of line EP(3)3321, EP(3)0979, and EP(3)0834, leading to increase of the energy storage triglycerides.

Example 3: Identification of human homologous genes and proteins

The Drosophila genes and proteins encoded thereby with functions in the regulation of triglyceride metabolism were further analysed using the BLAST algorithm searching in publicly available sequence databases and mammalian homologs were identified (see FIGURES 3, 4, 8, 9, 13, 14, 18, 19, 23, 24, 28, 29, 32, and 33).

As shown in FIGURE 3A, the gene product of Drosophila gilgamesh (gish; Gadfly Accession Number CG6963; Genbank Accession Number NM_080202) is 83% homologous over 426 amino acids (of 447 amino acids) to a human casein kinase 1 (also referred to as casein kinase 1, gamma 3) (GenBank Accession Number XM_049422 for the cDNA, XP_049422 for the protein). The gene product of Drosophila CG6963 is 80% homologous over 426 amino acids (of 459 amino acids) to human sequence 4 from patent WO0164905 (GenBank Accession Number AX239864). Casein kinase 1 homologous proteins and nucleic acid molecules coding therefore are obtainable from insect or vertebrate species, e.g. mammals or birds. Particularly preferred are human casein kinase 1 homologous nucleic acids and polypeptides encoded thereby, particularly encoding (i) human casein kinase 1, gamma 1 (Genbank Accession Numbers NM_022048, AB042563 for the cDNA, NP_071331 for the protein, Swiss Prot. Accession Number Q9HCP0 for the protein; see

Figure 3B and 3C; SEQ ID NO: 1 and 2), (ii) human casein kinase 1, gamma 2 (Genbank Accession Number NM_001319 for the cDNA, NP_001310 for the protein; see Figure 3D and 3E, SEQ ID NO: 3 and 4), or (iii) human casein kinase 1, gamma 3 (Genbank Accession Number NM_004384 for the cDNA, NP_003475 for the protein, formerly Genbank Accession Number XM_049422; see Figure 3F and 3G; SEQ ID NO: 5 and 6). An alignment of the casein kinases 1 from different species has been done by the Clustal W program and is illustrated in Figure 4.

As shown in FIGURE 8A, the gene product of Drosophila CG1534 (also referred to as Gadfly Accession Number CG32672) is 96% homologous over 113 amino acids to human GABARAP (Genbank Accession Number NP_009209.1), and to mouse GABARAP (Genbank Accession Number NP_062723.1). GABARAP homologous proteins and nucleic acid molecules coding therefore are obtainable from insect or vertebrate species, e.g. mammals or birds. Particularly preferred are human GABARAP homologous nucleic acids and polypeptides encoded thereby, particularly encoding (i) human GABARAP (Genbank Accession Number NM_007278 for the cDNA, NP_009209 for the protein; see Figure 8B and 8C; SEQ ID NO: 7 and 8), (ii) human GABARAP like 1 (Genbank Accession Number NM_031412 for the cCNA, NP_113600 for the protein; see Figure 8D and 8E; SEQ ID NO: 9 and 10), (iii) human GABARAP like 2 (Genbank Accession Number NM_007285 for the cDNA, NP_009216 for the protein; see Figure 8F and 8G; SEQ ID NO: 11 and 12), or (iv) human GABARAP like 3 (Genbank Accession Number NM_032568 for the cDNA, NP_115957 for the protein; see Figure 8H and 8I; SEQ ID NO: 13 and 14). An alignment of GABARAP and GABARAP like proteins from different species has been done by the Clustal W program and is illustrated in Figure 9.

As shown in FIGURE 13A, the gene product of Drosophila CG10576 is 70% homologous over 276 amino acids (of 386 amino acids) to human proliferation-associated 2G4, 38kD (also referred to as PA2G4, HG4-1, and

cell cycle protein; GenBank Accession Number XM_049048; Lamartine et al., 1997, Cytogenet. Cell Genet. 78:31-35), which is identical to sequence 5 from patent US 5,871,973 (sequence 5, GenBank Accession Number AAE06380.1). The gene product of Drosophila CG10576 is 70% homologous over 276 amino acids (of 386 amino acids) to mouse proliferation-associated 2G4, 38kD (GenBank Accession Number NM_011119), which is identical to sequence 10 from patent US 5,871,973 (GenBank Accession Number AAE06384.1). PA2G4 homologous proteins and nucleic acid molecules coding therefore are obtainable from insect or vertebrate species, e.g. mammals or birds. Particularly preferred are human PA2G4 homologous nucleic acids and polypeptides encoded thereby, particularly encoding human proliferation-associated 2G4 protein (Genbank Accession Number NM_006191 for the cDNA, NP_006182 for the protein, formerly Genbank Accession Number XM_049048; see Figure 13B and 13C; SEQ ID NO: 15 and 16). An alignment of PA2G4, 38 kDa homologs from different species has been done by the Clustal W program and is illustrated in Figure 14.

As shown in FIGURE 18A, the gene product of Drosophila Mocs1 (Gadfly Accession Number CG7858) is 77% homologous over 351 amino acids (of 385 amino acids) to human molybdenum cofactor biosynthesis protein A (also referred to as MOCSA; GenBank Accession Number AAB87523), and 77% homologous over 348 amino acids (of 385 amino acids) to human molybdenum cofactor synthesis 1 (also referred to as MOCS1; GenBank Accession Number XP_046687). Mocs1 homologous proteins and nucleic acid molecules coding therefore are obtainable from insect or vertebrate species, e.g. mammals or birds. Particularly preferred are human Mocs1 homologous nucleic acids and polypeptides encoded thereby, particularly encoding the human MOCS1 isoforms, for example (i) a human molybdenum cofactor biosynthesis protein A or a molybdenum cofactor biosynthesis protein C (also referred to as MOCSA or MOCSC; Genbank Accession Number AF034374 for the cDNA, AAB87523 for the protein;

see Figure 18B and 18C; SEQ ID NO: 17 and 18), (ii) human molybdenum cofactor synthesis 1 protein isoforms (MOCS1; Genbank Accession Numbers XM_046687, XM_166358, NM_005943, NM_005942, NM_138928 for the cDNAs, XP_046687, XP_166358, NP_005934, NP_005933, NP_620306 for the proteins; see Figures 18D, 18E, 18F, 18G, 18H, and 18I; SEQ ID NO: 19, 20, 21, 22, 23, and 24). An alignment of Mocs1 homologs from different species has been done by the Clustal W program and is illustrated in Figure 19.

As shown in FIGURE 23A, gene product of Drosophila peanut (pnut; Gadfly Accession Number CG8705) is 78% homologous over 331 amino acids (of 418 amino acids) to human cell division cycle 10 homolog (GenBank Accession Number NM_001788; Nakatsuru et al., 1994, BBR comm. 202:82-87), and 77% homologous over 302 amino acids (of 384 amino acids) to human CDC10 protein homolog, similar to septin 7 (GenBank Accession Number XM_011595). The gene product of Drosophila peanut is 78% homologous over 330 amino acids (of 417 amino acids) to mouse septin 7 (cell division cycle 10 homolog) (GenBank Accession Number AJ223782), and 78% homologous over 331 amino acids (of 419 amino acids) to Candida albicans septin 7 protein (GenBank Accession Number AAE20750.1, sequence 5 from patent US 5,849,556 and US 5,952,214). Peanut homologous proteins and nucleic acid molecules coding therefore are obtainable from insect or vertebrate species, e.g. mammals or birds. Particularly preferred are human peanut homologous nucleic acids and polypeptides encoded thereby, particularly encoding human cell division cycle 10 protein (CDC10; Genbank Accession Number XM_165879, NM_001788 for the cDNA, XP_165879, NP_001779 for the protein; formerly Genbank Accession Number XM_011595); see Figures 23B and 23C; SEQ ID NO: 25 and 26). An alignment of CDC10 homologs from different species has been done by the Clustal W program and is illustrated in Figure 24.

As shown in FIGURE 28A, gene product of Drosophila GadFly Accession Number CG7069 is 68% homologous over 412 amino acids (of 531 amino acids) to human pyruvate kinase, muscle (GenBank Accession Number XM_037768). The gene product of GadFly Accession Number CG7069 is 68% homologous over 416 amino acids (of 531 amino acids) to mouse pyruvate kinase 3 (GenBank Accession Number BC016619). Pyruvate kinase homologous proteins and nucleic acid molecules coding therefore are obtainable from insect or vertebrate species, e.g. mammals or birds. Particularly preferred are human pyruvate kinase homologous nucleic acids and polypeptides encoded thereby, particularly encoding (i) human human pyruvate kinase, muscle (also referred to as PKM1 and PKM2; Genbank Accession Number X56494 for the cDNA, P14618 and P14786 for the proteins; formerly Genbank Accession Number XM_037768; see Figure 28B, 28C, and 28D; SEQ ID NO: 27, 28, and 29), or (ii) human pyruvate kinase, liver and RBC (Genbank Accession Number NM_000298 for the cDNA, NP_000289 for the protein; see Figure 28E and 28F; SEQ ID NO: 30 and 31); see Figures 23B and 23C; SEQ ID NO: 25 and 26). An alignment of pyruvate kinase homologs from different species has been done by the Clustal W program and is illustrated in Figure 29.

As shown in FIGURE 32A, gene product of Drosophila calreticulin (Crc; Gadfly Accession Number CG9429) is 77% homologous over 404 amino acids (of 417 amino acids) to human calreticulin precursor (GenBank Accession Number NP_004334). Calreticulin homologous proteins and nucleic acid molecules coding therefore are obtainable from insect or vertebrate species, e.g. mammals or birds. Particularly preferred are human calreticulin homologous nucleic acids and polypeptides encoded thereby, particularly encoding (i) human calreticulin (Genbank Accession Numbers NM_004343, M84739 for the cDNA, NP_004334 for the protein; see Figure 32B and 32C; SEQ ID NO: 32 and 33), or (ii) human calreticulin 2 (hypothetical protein MGC26577; Genbank Accession Number NM_145046 for the cDNA, NP_659483 for the protein; see Figure 32D

and 32E; SEQ ID NO: 34 and 35). An alignment of calreticulin homologs from different species has been done by the Clustal W program and is illustrated in Figure 33.

Example 4: dUCPy modifier screen

Expression of Drosophila uncoupling protein dUCPy in a non-vital organ like the eye (Gal4 under control of the eye-specific promoter of the "eyeless" gene) results in flies with visibly damaged eyes. This easily visible eye phenotype is the basis of a genetic screen for gene products that can modify UCP activity.

Parts of the genomes of the strain with Gal4 expression in the eye and the strain carrying the pUAST-dUCPy construct were combined on one chromosome using genomic recombination. The resulting fly strain has eyes that are permanently damaged by dUCPy expression. Flies of this strain were crossed with flies of a large collection of mutagenized fly strains. In this mutant collection a special expression system (EP-element, Ref.: Rørth P, Proc Natl Acad Sci U S A 1996, 93(22):12418-22) is integrated randomly in different genomic loci. The yeast transcription factor Gal4 can bind to the EP-element and activate the transcription of endogenous genes close the integration site of the EP-element. The activation of the genes therefore occurs in the same cells (eye) that overexpress dUCPy. Since the mutant collection contains several thousand strains with different integration sites of the EP-element it is possible to test a large number of genes whether their expression interacts with dUCPy activity. In case a gene acts as an enhancer of UCP activity the eye defect will be worsened; a suppressor will ameliorate the defect.

Using this screen a gene with suppressing activity was discovered that was found to be the calreticulin gene in Drosophila.

Example 5: Genetic adipose pathway screen

Adipose (adp) is a protein that has been described as regulating, causing or contributing to obesity in an animal or human (see WO 01/96371). Transgenic flies containing a wild type copy of the adipose cDNA under the control of the Gal4/UAS system were generated (Brand and Perrimon, 1993, Development 118:401-415; for adipose cDNA, see WO 01/96371). Chromosomal recombination of these transgenic flies with an eyeless-Gal4 driver line has been used to generate a stable recombinant fly line over-expressing adipose in the developing Drosophila eye. Animals receiving transgenic adipose activity under these conditions developed into adult flies with a visible change of eye phenotype. Virgins of the recombinant driver line were crossed with males of the mutant EP-line collection in single crosses and kept for preferably 12 to 15 days at 29°C. The offspring was checked for modifications of the eye phenotype (enhancement or suppression). Mutations changing the eye phenotype affect genes that modify adipose activity. The inventors have found that the fly line HD-EP(3)37409 is an enhancer of the eye-adp-Gal4 induced phenotype. This result is strongly suggesting an interaction of gilgamesh gene with adipose since the integration of HD-EP(3)37409 was found to be located at the gilgamesh locus. This is supporting the function of gilgamesh and homologous proteins in the regulation of the energy homeostasis.

Example 6: Expression profiling experiments

To analyze the expression of the polypeptides disclosed in this invention in mammalian tissues, several mouse strains (preferrably mouse strains C57Bl/6J, C57Bl/6 ob/ob and C57Bl/KS db/db which are standard model systems in obesity and diabetes research) were purchased from Harlan Winkelmann (33178 Borchen, Germany) and maintained under constant temperature (preferrably 22°C), 40 per cent humidity and a light / dark cycle of preferrably 14 / 10 hours. The mice were fed a standard diet (for example, from ssniff Spezialitäten GmbH, order number ssniff M-Z V1126-000). For the fasting experiment ("fasted-mice"), wild type mice were starved for 48 h without food, but only water supplied ad libitum (see, for example, Schnetzler et al. J Clin Invest 1993 Jul;92(1):272-80, Mizuno et al. Proc Natl Acad Sci U S A 1996 Apr 16;93(8):3434-8). Animals were sacrificed at an age of 6 to 8 weeks. The animal tissues were isolated according to standard procedures known to those skilled in the art, snap frozen in liquid nitrogen and stored at -80°C until needed.

For analyzing the role of the proteins disclosed in this invention in the in vitro differentiation of different mammalian cell culture cells for the conversion of pre-adipocytes to adipocytes, mammalian fibroblast (3T3-L1) cells (e.g., Green & Kehinde, Cell 1: 113-116, 1974) were obtained from the American Tissue Culture Collection (ATCC, Hanassas, VA, USA; ATCC- CL 173). 3T3-L1 cells were maintained as fibroblasts and differentiated into adipocytes as described in the prior art (e.g., Qiu. et al., J. Biol. Chem. 276:11988-95, 2001; Slieker et al., BBRC 251: 225-9, 1998). At various time points of the differentiation procedure, beginning with day 0 (day of confluence) and day 2 (hormone addition; for example, dexamethasone and 3-isobutyl-1-methylxanthine), up to 10 days of differentiation, suitable aliquots of cells were taken every two days. Alternatively, mammalian fibroblast 3T3-F442A cells (e.g., Green & Kehinde, Cell 7: 105-113, 1976) were obtained from the Harvard Medical School, Department of Cell Biology (Boston, MA, USA). 3T3-F442A cells were maintained as fibroblasts and differentiated into adipocytes as

described previously (Djian, P. et al., J. Cell. Physiol., 124:554-556, 1985). At various time points of the differentiation procedure, beginning with day 0 (day of confluence and hormone addition, for example, insulin), up to 10 days of differentiation, suitable aliquots of cells were taken every two days. 3T3-F442A cells are differentiating in vitro already in the confluent stage after hormone (insulin) addition.

RNA was isolated from mouse tissues or cell culture cells using Trizol Reagent (for example, from Invitrogen, Karlsruhe, Germany) and further purified with the RNeasy Kit (for example, from Qiagen, Germany) in combination with an DNase-treatment according to the instructions of the manufacturers and as known to those skilled in the art. Total RNA was reverse transcribed (preferrably using Superscript II RNaseH- Reverse Transcriptase, from Invitrogen, Karlsruhe, Germany) and subjected to Taqman analysis preferrably using the Taqman 2xPCR Master Mix (from Applied Biosystems, Weiterstadt, Germany; the Mix contains according to the Manufacturer for example AmpliTaq Gold DNA Polymerase, AmpErase UNG, dNTPs with dUTP, passive reference Rox and optimized buffer components) on a GeneAmp 5700 Sequence Detection System (from Applied Biosystems, Weiterstadt, Germany).

Taqman analysis of casein kinase 1, gamma 1 (CK1G1) was performed preferrably using the following primer/probe pairs: mouse CK1G1 forward primer (Seq ID NO: 36) 5'- AAT GTC GAT GAC CCC ACT GG-3'; mouse CK1G1 reverse primer (Seq ID NO: 37) 5'- TCC ACT ACC TCC ACT TCG GC -3'; mouse CK1G1 Taqman probe (Seq ID NO: 38) (5/6-FAM) TCA CTC CAA TGC ACC AAT CAC AGC TCA (5/6-TAMRA).

Taqman analysis of casein kinase 1, gamma 3 (CK1G3) was performed preferrably using the following primer/probe pairs: mouse CK1G3 forward primer (Seq ID NO: 39) 5' AAA TGG AGA GCT GAA CAC GGA -3'; mouse CK1G3 reverse primer (Seq ID NO: 40) 5'- TGT AGG AGC TGT AAT GGG

TGC A -3'; mouse CK1G3 Taqman probe (Seq ID NO: 41) (5/6-FAM) CCC CAC GGC AGG ACG GTC G (5/6-TAMRA).

Taqman analysis of GABARAP 2 was performed preferrably using the following primer/probe pairs: mouse GABARAP 2 forward primer (Seq ID NO: 42) 5'- TCA GCC CAG GAA GAA CTT GTG-3'; mouse GABARAP 2 reverse primer (Seq ID NO: 43) 5'- CAA GGC TGT GAT TCA TGT CGT C -3'; mouse GABARAP 2 Taqman probe (Seq ID NO: 44) (5/6-FAM) TGC ATT GGC TGT GAG AGC GGG AT (5/6-TAMRA).

Taqman analysis of the proliferation associated protein 2G4, 38kDa (PA2G4) was performed preferrably using the following primer/probe pairs: mouse PA2G4 forward primer (Seq ID NO: 45) 5'- AGA CGA GCA GCA GGA GCA A -3'; mouse PA2G4 reverse primer (Seq ID NO: 46) 5'- TGT CGC CCC CCA TCT TAT AC -3'; mouse PA2G4 Taqman probe (Seq ID NO: 47) (5/6-FAM) ATC GCC GAG GAC CTG GTC GTG AC (5/6-TAMRA).

Taqman analysis of Mocs was performed preferrably using the following primer/probe pairs: mouse Mocs forward primer (Seq ID NO: 48) 5'- CCT GAG CCA CGT GCA GGT -3'; mouse Mocs reverse primer (Seq ID NO: 49) 5'- AGG ATG CCT GGA TCA ACA CAG -3'; mouse Mocs Taqman probe (Seq ID NO: 50) (5/6-FAM) CAC CTG GAG TTA GAC AGC ACA CGC CA (5/6-TAMRA).

Taqman analysis of the peanut homologous protein (Peanut) was performed preferrably using the following primer/probe pairs: mouse Peanut forward primer (Seq ID NO: 51) 5'- CGA GGA GAG GAG CGT CAA CT -3'; mouse Peanut reverse primer (Seq ID NO: 52) 5'- CCC ACA TAG CCC TCA AGG TTC -3'; mouse Peanut Taqman probe (Seq ID NO: 53) (5/6-FAM) CGG CAC CAT GGC TCA ACC GA (5/6-TAMRA).

Expression profiling studies confirm the particular relevance of casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, and CDC10 as regulators of energy metabolism in mammals. The results are shown in FIGURES 5, 10, 15, 20, and 25. casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, and CDC10 show expression in many tissues. In addition, significant expression levels of casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, and CDC10 were found in metabolic active tissues like white adipocyte tissue (WAT) and brown adipocyte tissue (BAT), (FIGURE 5A, 5C, 10A, 15A, 20A, and 25A), confirming a role in the regulation of energy homeostasis and thermogenesis.

Further, we show that casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, and CDC10 are regulated by fasting and by genetically induced obesity, and that thus the expression of casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, and CDC10 is under metabolic control. In this invention, we used mouse models of insulin resistance and/or diabetes, such as mice carrying gene knockouts in the leptin pathway (for example, ob (leptin) or db (leptin receptor/ligand) mice) to study the expression of the protein of the invention. Such mice develop typical symptoms of diabetes, show hepatic lipid accumulation and frequently have increased plasma lipid levels (see Bruning et al, 1998, Mol. Cell. 2:449-569).

The GABARAP protein was also examined in the in vitro differentiation models for the conversion of pre-adipocytes to adipocytes, as described above.

As shown in Figure 5A and 5C, real time PCR (Taqman) analysis of the expression of the casein kinase 1, gamma 1 (CK1G1) and casein kinase 1, gamma 3 (CK1G3) RNA in mammalian (mouse) tissues revealed that CK1G1 and CK1G3 are expressed in different mammalian tissues, including white adipose tissue (WAT), brown adipose tissue (BAT), hypothalamus, and brain. The high experession levels of CK1G1 and CK1G3 in these

tissues indicates, that CK1G1 and CK1G3 are involved in the metabolism of tissues relevant for the metabolic syndrome. The expression of CK1G1 and CK1G3 are under metabolic control: In genetically obese (ob/ob) mice, expression of CK1G1 and CK1G3 are strongly induced in BAT (see Figure 5B and 5D). Analysis of the expression of casein kinase 1 gamma 2 (CK1G2) revealed that CK1G2 is expressed in different mammalian tissues with strongest expression in testis, and a 2.5 fold higher expression of CK1G2 in brown adipose tissue of genetically obese (ob/ob) mice, compared to wild type mice (data not shown).

As shown in Figure 10A, real time PCR (Taqman) analysis of the expression of the GABARAP 2 RNA in mammalian (mouse) tissues revealed that GABARAP 2 is expressed in different mammalian tissues, including white adipose tissue (WAT), brown adipose tissue (BAT), liver, hypothalamus, brain, and kidney. The high experession levels of GABARAP 2 in these tissues indicates, that GABARAP 2 is involved in the metabolism of tissues relevant for the metabolic syndrome. The expression of GABARAP 2 is under metabolic control: In fasted mice, expression of GABARAP 2 is strongely induced in muscle (see Figure 10B). GABARAB-2 is down regulated during the clonal expansion phase of preadipocyte differentaition. It is present in the differentiated adipocyte (see Figure 10C).

As shown in Figure 15A, real time PCR (Taqman) analysis of the expression of the PA2G4 RNA in mammalian (mouse) tissues revealed that PA2G4 is expressed in different mammalian tissues, including white adipose tissue (WAT), brown adipose tissue (BAT), and brain. The high experession levels of PA2G4 in these tissues indicates, that PA2G4 is involved in the metabolism of tissues relevant for the metabolic syndrome. The expression of PA2G4 is under metabolic control: In genetically obese (ob/ob) mice, expression of PA2G4 is strongely induced in BAT and heart, and in fasted mice, expression of PA2G4 is strongly induced in heart (see Figure 15B).

As shown in Figure 20A, real time PCR (Taqman) analysis of the expression of the Mocs RNA in mammalian (mouse) tissues revealed that Mocs is rather ubiquitously expressed in wildtype mice. The expression of Mocs in brown adipose tissue is under metabolic control: In genetically obese (ob/ob) mice, expression is strongely induced compared to wildtype levels (see Figure 20B).

As shown in Figure 25A, real time PCR (Taqman) analysis of the expression of the the Peanut homologous RNA in mammalian (mouse) tissues revealed that the Peanut homolog is rather ubiquitously expressed in wildtype mice. The expression of the Peanut homolog in brown adipose tissue is under metabolic control: In genetically obese (ob/ob) mice, expression is strongely induced compared to wildtype levels (see Figure 25B).

Example 6: In vitro assays for the determination of triglyceride storage, synthesis and transport

Obesity is known to be caused by different reasons such as non-insulin dependent diabetes, increase in triglycerides, increase in carbohydrate bound energy and low energy expenditure. For example, an increase in energy expenditure (and thus, lowering the body weight) would include the elevated utilization of both circulating and intracellular glucose and triglycerides, free or stored as glycogen or lipids as fuel for energy and/or heat production. The cellular level of triglycerides and glycogen is analyzed in cells overexpressing the protein of the invention.

Preparation of cell lysates for analysis of metabolites

Starting at confluence (d0), cell media was changed every 48 hours. Cells and media were harvested 8 hours prior to media change as follows. Media

was collected, and cells were washed twice in PBS prior to lyses in 600 µl HB-buffer (0.5% polyoxyethylene 10 tridecylethane, 1 mM EDTA, 0.01M NaH_2PO_4, pH 7.4). After inactivation at 70°C for 5 minutes, cell lysates were prepared on Bio 101 systems lysing matrix B (0.1 mm silica beads; Q-Biogene, Carlsbad, USA) by agitation for 2 x 45 seconds at a speed of 4.5 (Fastprep FP120, Bio 101 Thermosavant, Holbrock, USA). Supernatants of lysed cells were collected after centrifugation at 3000 rpm for 2 minutes, and stored in aliquots for later analysis at -80°C.

Changes in cellular triglyceride levels during adipogenesis

Cell lysates and media were simultaneously analysed in 96-well plates for total protein and triglyceride content using the Bio-Rad DC Protein assay reagent (Bio-Rad, Munich, Germany) according to the manufacturer's instructions and a modified enzymatic triglyceride kit (GPO-Trinder; Sigma) briefly final volumes of reagents were adjusted to the 96-well format as follows: 10 µl sample was incubated with 200 µl reagent A for 5 minutes at 37°C. After determination of glycerol (initial absorbance at 540 nm), 50 µl reagent B was added followed by another incubation for 5 minutes at 37°C (final absorbance at 540 nm). Glycerol and triglyceride concentrations were calculated using a glycerol standard set (Sigma) for the standard curve included in each assay.

Changes in cellular glycogen levels during adipogenesis

Cell lysates and media were simultaneously analysed in triplicates in 96-well plates for total protein and glycogen content using the Bio-Rad DC Protein assay reagent (Bio-Rad, Munich, Germany) according to the manufacturer's instructions and an enzymatic starch kit from Hoffmann-La Roche (Basel, Switzerland). 10-µl samples were incubated with 20-µl amyloglucosidase solution for 15 minutes at 60°C to digest glycogen to glucose. The glucose is further metabolised with 100 µl distilled water and 100 µl of enzyme cofactor buffer and 12 µl of enzyme buffer (hexokinase and glucose phosphate dehydrogenase). Background glucose levels are

determined by subtracting values from a duplicate plate without the amyloglucosidase. Final absorbance is determined at 340 nm. HB-buffer as blank, and a standard curve of glycogen (Hoffmann-La Roche) were included in each assay. Glycogen contents in samples were calculated using a standard curve.

Synthesis of lipids during adipogenesis

During the terminal stage of adipogenesis (day 12) cells were analysed for their ability to metabolise lipids. A modified protocol to the method of Jensen et al (2000) for lipid synthesis was established. Cells were washed 3 times with PBS prior to serum starvation in Krebs-Ringer-Bicarbonate-Hepes buffer (KRBH; 134 nM NaCl, 3.5 mM KCl, 1.2 mM KH_2PO_4, 0.5 mM $MgSO_4$, 1.5 mM $CaCl_2$, 5 mM $NaHCO_3$, 10 mM Hepes, pH 7.4), supplemented with 0.1% FCS for 2.5h at 37°C. For insulin-stimulated lipid synthesis, cells were incubated with 1 μM bovine insulin (Sigma; carrier: 0.005N HCl) for 45min at 37°C. Basal lipid synthesis was determined with carrier only. ^{14}C(U)-D-glucose (NEN Life Sciences) in a final activity of 1μCi/Well/ml in the presence of 5 mM glucose was added for 30 min at 37°C. For the calculation of background radioactivity, 25μM cytochalasin B (Sigma) was used. All assays were performed in duplicate wells. To terminate the reaction, cells were washed 3 times with ice cold PBS, and lysed in 1 ml 0.1N NaOH. Protein concentration of each well was assessed using the standard Biuret method (Protein assay reagent; Bio-Rad). Total lipids were separated from aqueous phase after overnight extraction in Insta-Fluor scintillation cocktail (Packard Bioscience) followed by scintillation counting.

Transport and metabolism of free fatty acids during adipogenesis

During the terminal stage of adipogenesis (d12) cells were analysed for their ability to transport long chain fatty acid across the plasma membrane. A modified protocol to the method of Abumrad et al (1991) (Proc. Natl. Acad. Sci. USA, 1991: 88; 6008-12) for cellular transportation of fatty

acid was established. In summary, cells were washed 3 times with PBS prior to serum starvation. This was followed by incubation in KRBH buffer supplemented with 0.1% FCS for 2.5h at 37°C. Uptake of exogenous free fatty acids was initiated by the addition of isotopic media containing non radioactive oleate and (^{3}H)oleate (NEN Life Sciences) complexed to serum albumin in a final activity of 1μCi/Well/ml in the presence of 5 mM glucose for 30min at room temperature (RT). For the calculation of passive diffusion (PD) in the absence of active transport (AT) across the plasma membrane 20mM of phloretin in glucose free media (Sigma) was added for 30 min at RT. All assays were performed in duplicate wells. To terminate the active transport 20mM of phloretin in glucose free media was added to the cells. Cells were lysed in 1 ml 0.1N NaOH and the protein concentration of each well were assessed using the standard Biuret method (Protein assay reagent; Bio-Rad). Esterified fatty acids were separated from free fatty acids using overnight extraction in Insta-Fluor scintillation cocktail (Packard Bioscience) followed by scintillation counting.

Example 7: Glucose uptake assay

For the determination of glucose uptake, cells were washed 3 times with PBS prior to serum starvation in KRBH buffer supplemented with 0.1% FCS and 0.5mM Glucose for 2.5h at 37°C. For insulin-stimulated glucose uptake, cells were incubated with 1 μM bovine insulin (Sigma; carrier: 0.005N HCl) for 45 min at 37°C. Basal glucose uptake was determined with carrier only. Non-metabolizable 2-deoxy-3H-D-glucose (NEN Life Science, Boston, USA) in a final activity of 0,4 μCi/Well/ml was added for 30 min at 37°C. For the calculation of background radioactivity, 25 μM cytochalasin B (Sigma) was used. All assays were performed in duplicate wells. To terminate the reaction, cells were washed 3 times with ice cold PBS, and lysed in 1 ml 0.1N NaOH. Protein concentration of each well was assessed using the standard Biuret method (Protein assay reagent; Bio-Rad), and scintillation counting of cell lysates in 10 volumes

Ultima-gold cocktail (Packard Bioscience, Groningen, Netherlands) was performed.

Example 8: Generation and analysis of transgenic mice

Generation of the transgenic animals

Mouse cDNA was isolated from mouse brown adipose tissue (BAT) using standard protocols as known to those skilled in the art. The cDNA was amplified by RT-PCR and point mutations were introduced into the cDNA.

The resulting mutated cDNA was cloned into a suitable transgenic expression vector. The transgene was microinjected into the male pronucleus of fertilized mouse embryos (preferably strain C57/BL6/CBA F1 (Harlan Winkelmann). Injected embryos were transferred into pseudo-pregnant foster mice. Transgenic founders were detected by PCR analysis. Two independent transgenic mouse lines containing the construct were established and kept on a C57/BL6 background. Briefly, founder animals were backcrossed with C57/BL6 mice to generate F1 mice for analysis. Transgenic mice were continously bred onto the C57/Bl6 background. The expression of the proteins of the invention can be analyzed by taqman analysis as described above, and further analysis of the mice can be done as known to those skilled in the art.

All publications and patents mentioned in the above specification are herein incorporated by reference.

Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly

limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

5

Claims

1. A pharmaceutical composition comprising a nucleic acid molecule of the casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin gene family or a polypeptide encoded thereby or a fragment or a variant of said nucleic acid molecule or said polypeptide or an antibody, an aptamer or another receptor recognizing a nucleic acid molecule of the casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin gene family or a polypeptide encoded thereby together with pharmaceutically acceptable carriers, diluents and/or adjuvants.

2. The composition of claim 1, wherein the nucleic acid molecule is a vertebrate or insect casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin nucleic acid, particularly encoding human casein kinase 1, gamma 1 (SEQ ID NO: 1), human casein kinase 1, gamma 2 (SEQ ID NO: 3), human casein kinase 1, gamma 3 (SEQ ID NO: 5), human GABARAP (SEQ ID NO: 7), human GABARAP like 1 (SEQ ID NO: 9), human GABARAP like 2 (SEQ ID NO: 11), human GABARAP like 3 (SEQ ID NO: 13), human PA2G4 (SEQ ID NO: 15), human MOCSA (SEQ ID NO: 17), human MOCS1 isoform 1 (SEQ ID NO: 19), human MOCS1 isoform 2 (SEQ ID NO: 21), human MOCS1 isoform 3 (SEQ ID NO: 23), human CDC10 (SEQ ID NO: 25), human pyruvate kinase, muscle (SEQ ID NO: 27), human pyruvate kinase, liver and RBC (SEQ ID NO: 30), human calreticulin (SEQ ID NO: 32), human calreticulin 2 (SEQ ID NO:34), or a fragment there of or a variant thereof.

3. The composition of claim 1 or 2, wherein said nucleic acid molecule

(a) hybridizes at 50°C in a solution containing 1 x SSC and 0.1% SDS to a nucleic acid molecule as defined in claim 2 or a nucleic acid molecule which is complementary thereto;

(b) it is degenerate with respect to the nucleic acid molecule of (a)

(c) encodes a polypeptide which is at least 85%, preferably at least 90%, more preferably at least 95%, more preferably at least 98% and up to 99,6% identical to casein kinase 1, gamma 1 (SEQ ID NO: 2), human casein kinase 1, gamma 2 (SEQ ID NO: 4), human casein kinase 1, gamma 3 (SEQ ID NO: 6), human GABARAP (SEQ ID NO: 8), human GABARAP like 1 (SEQ ID NO: 10), human GABARAP like 2 (SEQ ID NO: 12), human GABARAP like 3 (SEQ ID NO: 14), human PA2G4 (SEQ ID NO: 16), human MOCSA (SEQ ID NO: 18), human MOCS1 isoform 1 (SEQ ID NO: 20), human MOCS1 isoform 2 (SEQ ID NO: 22), human MOCS1 isoform 3 (SEQ ID NO: 24), human CDC10 (SEQ ID NO: 26), human pyruvate kinase, muscle, isozyme M1 (SEQ ID NO: 28), human pyruvate kinase, muscle, isozyme M2 (SEQ ID NO: 29), human pyruvate kinase, liver and RBC (SEQ ID NO: 31), human calreticulin (SEQ ID NO: 33), human calreticulin 2 (SEQ ID NO:35), as defined in claim 2;

(d) differs from the nucleic acid molecule of (a) to (c) by mutation and wherein said mutation causes an alteration, deletion, duplication or premature stop in the encoded polypeptide.

4. The composition of any one of claims 1-3, wherein the nucleic acid molecule is a DNA molecule, particularly a cDNA or a genomic DNA.

5. The composition of any one of claims 1-4, wherein said nucleic acid encodes a polypeptide contributing to regulating the energy homeostasis and/or the metabolism of triglycerides.

6. The composition of any one of claims 1-5, wherein said nucleic acid molecule is a recombinant nucleic acid molecule.

7. The composition of any one of claims 1-6, wherein the nucleic acid molecule is a vector, particularly an expression vector.

8. The composition of any one of claims 1-5, wherein the polypeptide is a recombinant polypeptide.

9. The composition of claim 8, wherein said recombinant polypeptide is a fusion polypeptide.

10. The composition of any one of claims 1-7, wherein said nucleic acid molecule is selected from hybridization probes, primers and anti-sense oligonucleotides.

11. The composition of any one of claims 1-10 which is a diagnostic composition.

12. The composition of any one of claims 1-10 which is a therapeutic composition.

13. The composition of any one of claims 1-12 for the manufacture of an agent for detecting and/or verifying, for the treatment, alleviation and/or prevention of an disorders, including metabolic diseases such as obesity and other body-weight regulation disorders as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hyper-

cholesterolemia, dyslipidemia, osteoarthritis, gallstones, and others, in cells, cell masses, organs and/or subjects.

14. Use of a nucleic acid molecule of the casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin gene family or a polypeptide encoded thereby or a fragment or a variant of said nucleic acid molecule or said polypeptide or an antibody, an aptamer or another receptor recognizing a nucleic acid molecule of the casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin gene family or a polypeptide encoded thereby for controlling the function of a gene and/or a gene product which is influenced and/or modified by a casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin homologous polypeptide.

15. Use of the nucleic acid molecule of the casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin gene family or a polypeptide encoded thereby or a fragment or a variant of said nucleic acid molecule or said polypeptide or an antibody, an aptamer or another receptor recognizing a nucleic acid molecule of the casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin gene family or a polypeptide encoded thereby for identifying substances capable of interacting with a casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin homologous polypeptide.

16. A non-human transgenic animal exhibiting a modified expression of a casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin homologous polypeptide.

17. The animal of claim 16, wherein the expression of the casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin homologous polypeptide is increased and/or reduced.

18. A recombinant host cell exhibiting a modified expression of a casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin homologous polypeptide.

19. The cell of claim 18 which is a human cell.

20. A method of identifying a (poly)peptide involved in the regulation of energy homeostasis and/or metabolism of triglycerides in a mammal comprising the steps of

(a) contacting a collection of (poly)peptides with a casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin homologous polypeptide or a fragment thereof under conditions that allow binding of said (poly)peptides;

(b) removing (poly)peptides which do not bind and

(c) identifying (poly)peptides that bind to said casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin homologous polypeptide.

21. A method of screening for an agent which modulates the interaction of a casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin homologous polypeptide with a binding target/agent, comprising the steps of

(a) incubating a mixture comprising

(aa) a casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin homologous polypeptide, or a fragment thereof;

(ab) a binding target/agent of said casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin homologous polypeptide or fragment thereof; and

(ac) a candidate agent

under conditions whereby said casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin polypeptide or fragment thereof specifically binds to said binding target/agent at a reference affinity;

(b) detecting the binding affinity of said casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin polypeptide or fragment thereof to said binding target to determine an (candidate) agent-biased affinity; and

(c) determining a difference between (candidate) agent-biased affinity and the reference affinity.

22. A method of screening for an agent which modulates the activity of a casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin homologous polypeptide comprising the steps of

(a) incubating a mixture comprising

(aa) a casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin homologous polypeptide, or a fragment thereof, and

(ab) a candidate agent

under conditions whereby said casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin polypeptide or fragment thereof has a reference activity;

(b) detecting the activity of said casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin polypeptide or fragment thereof to determine an (candidate) agent-biased activity and

(c) determining a difference between (candidate) agent-biased activity and reference activity.

23. A method of producing a composition comprising the (poly)peptide identified by the method of claim 20 or the agent identified by the method of claim 21 or 22 with a pharmaceutically acceptable carrier, diluent and/or adjuvant.

24. The method of claim 23 wherein said composition is a pharmaceutical composition for preventing, alleviating or treating of diseases and disorders, including metabolic diseases such as obesity and other body-weight regulation disorders as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, gallstones, and other diseases and disorders.

25. Use of a (poly)peptide as identified by the method of claim 20 or of an agent as identified by the method of claim 21 or 22 for the preparation of a pharmaceutical composition for the treatment, alleviation and/or prevention of diseases and disorders, including metabolic diseases such as obesity and other body-weight regulation disorders as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, gallstones, and other diseases and disorders.

26. Use of a nucleic acid molecule of the casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin family or of a fragment thereof for the preparation of a non-human animal which over- or under-expresses the casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin gene product.

27. Kit comprising at least one of

(a) a casein kinase 1 gamma, GABARAP, PA2G4, MOCS1, CDC10, PK, or calreticulin nucleic acid molecule or a fragment thereof;

(b) a vector comprising the nucleic acid of (a);

(c) a host cell comprising the nucleic acid of (a) or the vector of (b);

(d) a polypeptide encoded by the nucleic acid of (a);

(e) a fusion polypeptide encoded by the nucleic acid of (a);

(f) an antibody, an aptamer or another receptor against the nucleic acid of (a) or the polypeptide of (d) or (e) and

(g) an anti-sense oligonucleotide of the nucleic acid of (a).

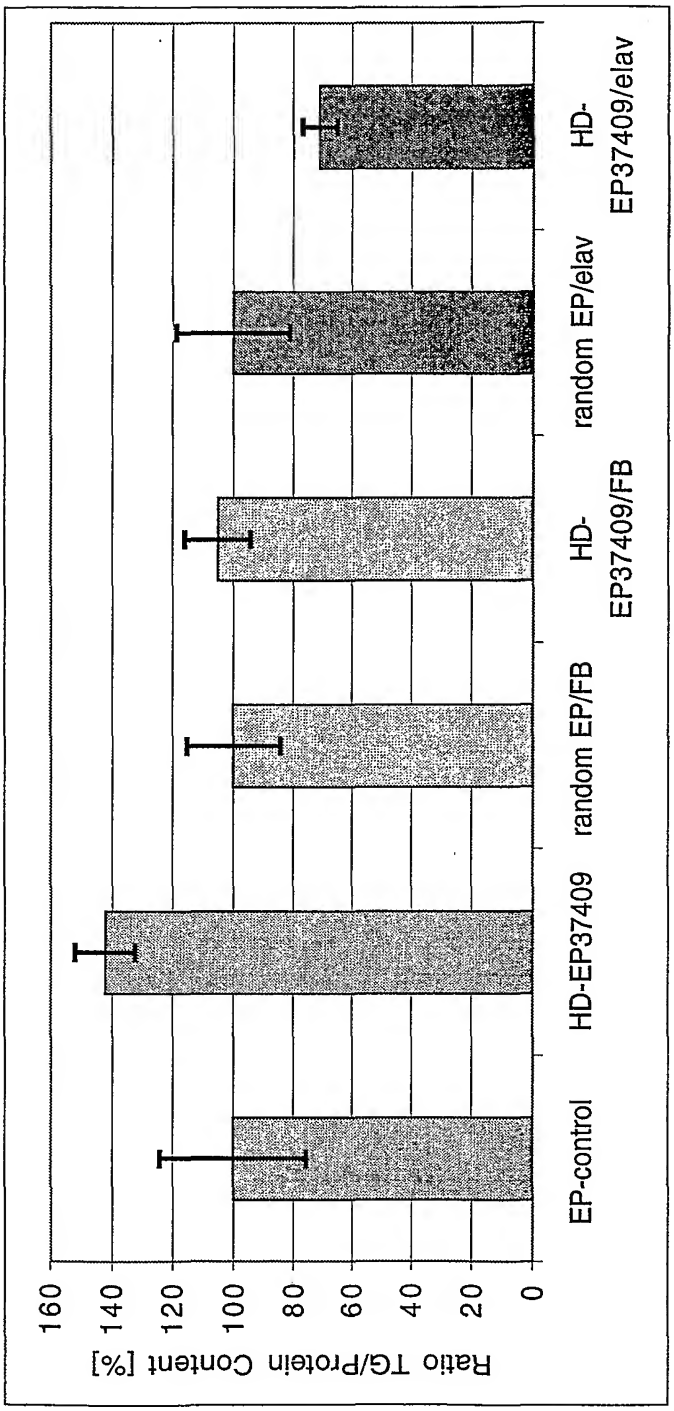


FIGURE 1. Triglyceride content of a gilgamesh (gish; Gadfly Accession Number CG6963) mutant

FIGURE 2. Molecular organisation of the casein kinase I gene *gilgamesh* (GadFly Accession Number CG6963)

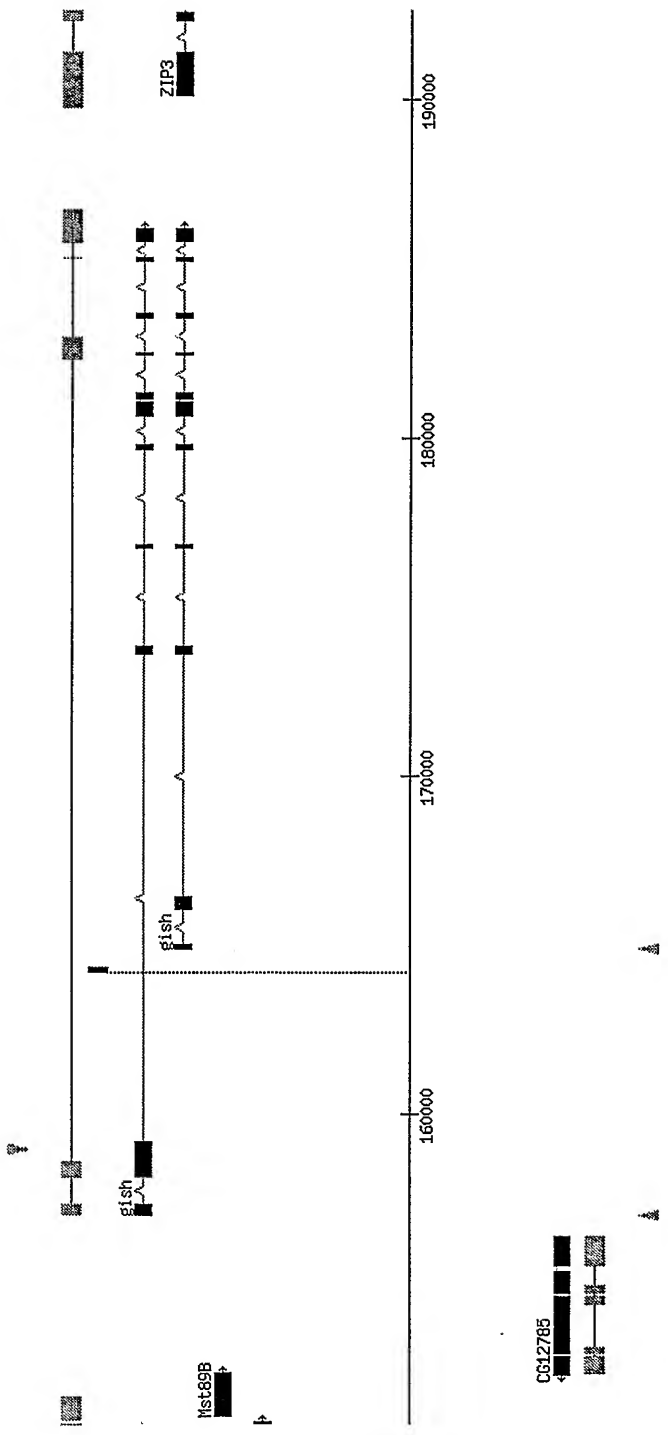


FIGURE 3: HUMAN HOMOLOG OF CG6963 (gilgamesh)

FIGURE 3A. BLASTP result for gilgamesh (Gadfly Accession Number CG6963)

Homology to human gene ref XM_049422; protein ref XP_049422.2

```
>ref|XP_049422.2| (XM_049422) casein kinase 1, gamma 3 [Homo sapiens]
Length = 447

Score =  628 bits (1601), Expect = e-178
Identities = 307/426 (72%), Positives = 356/426 (83%), Gaps = 10/426 (2%)

Query: 2   YSTRQSVSTTTGVLMVGPNFRVGKKIGCGNFGELRLGKNLYNNEHVAIKMEPMKSKAPQL 61
           ++TR + S+++GVLMVGPNFRVGKKIGCGNFGELRLGKNLY NE+VAIK+EPMKS+APQL
Sbjct: 24  HNTRGTGSSSSGVLMVGPNFRVGKKIGCGNFGELRLGKNLYTNEYVAIKLEPMKSRAPQL 83

Query: 62  HLEYRFYKLLGSHAEGVPEVYYFGPCGKYNALVMELLGPSLEDLFDICGRRFTLKSVLLI 121
           HLEYRFYK LGS  +G+P+VYYFGPCGKYNA+V+ELLGPSLEDLFD+C R F+LK+VL+I
Sbjct: 84  HLEYRFYKQLGS-GDGIPQVYYFGPCGKYNAMVLELLGPSLEDLFDLCDRTFSLKTVLMI 142

Query: 122 AIQLLHRIEYVHSRHLIYRDVKPENFLIGRTSTKREKIIHIIDFGLAKEYIDLDTNRHIP 181
           AIQL+ R+EYVHS++LIYRDVKPENFLIGR   K +++IHIIDFGLAKEYID +T +HIP
Sbjct: 143 AIQLISRMEYVHSKNLIYRDVKPENFLIGRPGNKTQQVIHIIDFGLAKEYIDPETKKHIP 202

Query: 182 YREHKSLTGTARYMSINTHMGREQSRRDDLEALGHMFMYFLRGSLPWQGLKADTLKERYQ 241
           YREHKSLTGTARYMSINTH+G+EQSRRDDLEALGHMFMYFLRGSLPWQGLKADTLKERYQ
Sbjct: 203 YREHKSLTGTARYMSINTHLGKEQSRRDDLEALGHMFMYFLRGSLPWQGLKADTLKERYQ 262

Query: 242 KIGDTKRATPIEVLCDGHPEEFATYLRYVRRLDFFETPDYDFLRRLFQDLFDRKGYTDEG 301
           KIGDTKRATPIEVLC+  P E ATYLRYVRRLDFFE PDYD+LR+LF DLFDRKGY  +
Sbjct: 263 KIGDTKRATPIEVLCENFP-EMATYLRYVRRLDFFEKPDYDYLRKLFTDLFDRKGYMFDY 321

Query: 302 EFDWTGKTMSTPVGSLQTGHEVIISPNKDRHN-----VTAKTNAKGGVAAWPDVPKPGAT 356
           E+DW GK + TPVG++Q   +  +S N++ H         +K  +    AAW
Sbjct: 322 EYDWIGKQLPTPVGAVQ--QDPALSSNREAHQHRDKMQQSKNQSADHRAAWDSQQANPHH 379

Query: 357 LGNLTPADRH-GSVQVVSSTNGELNPDDPTAGHSNTPITQQPEVEVVDETKCCCFFKRKK 415
           L     ADRH GSVQVVSSTNGELN DDPTAG SN PIT   EVEV+DETKCCCFFKR+K
Sbjct: 380 LRAHLAADRHGGSVQVVSSTNGELNTDDPTAGRSNAPITAPTEVEVMDETKCCCFFKRRK 439

Query: 416 KKSTRQ 421
           +K+ ++
Sbjct: 440 RKTIQR 445
```

FIGURE 3B: Predicted nucleotide sequence encoding human casein kinase 1, gamma 1 (SEQ ID NO:1)

```
   1 aactccttac ctttctctga ctacaattta tttggacata cttttgtatt gaagagaggt
  61 atacatactg aagctacttg ctgtactata ggagactctg tcctgtagga tcatggacca
 121 tcctagtagg gaaaaggatg aaagacaacg gacaactaaa cccatggcac aaaggagtgc
 181 acactgctct cgaccatctg gctcctcatc gtcctctggg gttcttatgg tgggacccaa
 241 cttcagggtt ggcaagaaga taggatgtgg gaacttcgga gagctcagat taggtaaaaa
 301 tctctacacc aatgaatatg tagcaatcaa actggaacca ataaaatcac gtgctccaca
 361 gcttcattta gagtacagat tttataaaca gcttggcagt gcaggtgaag gtctcccaca
 421 ggtgtattac tttggaccat gtgggaaata taatgccatg gtgctggagc tccttggccc
 481 tagcttggag gacttgtttg acctctgtga ccgaacattt actttgaaga cggtgttaat
 541 gatagccatc cagctgcttt ctcgaatgga atacgtgcac tcaaagaacc tcatttaccg
 601 agatgtcaag ccagagaact tcctgattgg tcgacaaggc aataagaaag agcatgttat
 661 acacattata gactttggac tggccaagga atacattgac cccgaaacca aaaaacacat
 721 accttatagg gaacacaaaa gtttaactgg aactgcaaga tatatgtcta tcaacacgca
 781 tcttggcaaa gagcaaagcc ggagagatga tttggaagcc ctaggccata tgttcatgta
 841 tttccttcga ggcagcctcc cctggcaagg actcaaggct gacacattaa aagagagata
 901 tcaaaaaatt ggtgacacca aaaggaatac tcccattgaa gctctctgtg agaactttcc
 961 agaggagatg gcaacctacc ttcgatatgt caggcgactg gacttctttg aaaaacctga
1021 ttatgagtat ttacggaccc tcttcacaga cctctttgaa aagaaaggct acacctttga
1081 ctatgcctat gattgggttg ggagacctat tcctactcca gtagggtcag ttcacgtaga
1141 ttctggtgca tctgcaataa ctcgagaaag ccacacacat agggatcggc catcacaaca
1201 gcagcctctt cgaaatcagg tggttagctc aaccaatgga gagctgaatg ttgatgatcc
1261 cacgggagcc cactccaatg caccaatcac agctcatgcc gaggtggagg tagtggagga
1321 agctaagtgc tgctgtttct ttaagaggaa aaggaagaag actgctcagc gccacaagtg
1381 accagtgcct cccaggagtc ctcaggccct ggggactctg actcaattgt acctgcagca
1441 tttctcattg gaaggggact cctctttggg ggagggtgga tatccaaacc aaaaagaaga
1501 aaacagatgc ccccagaagg gggccagtgc gggcagccag ggcctagtgg gtcattggcc
1561 atctccgctg ctaaggctct gagcaggtcc agagctgctg ttcctccact gcttgcccat
1621 agggctgcct ggttgactct cttccattg
```

FIGURE 3C: Predicted amino acid sequence of human casein kinase 1, gamma 1 (SEQ ID NO:2)

```
   1 mdhpsrekde rqrttkpmaq rsahcsrpsg sssssgvlmv gpnfrvgkki gcgnfgelrl
  61 gknlytneyv aiklepiksr apqlhleyrf ykqlgsageg lpqvyyfgpc gkynamvlel
 121 lgpsledlfd lcdrtftlkt vlmiaiqlls rmeyvhsknl iyrdvkpenf ligrqgnkke
 181 hvihiidfgl akeyidpetk khipyrehks ltgtarymsi nthlgkeqsr rddlealghm
 241 fmyflrgslp wqglkadtlk eryqkigdtk rntpiealce nfpeematyl ryvrrldffe
 301 kpdyeylrtl ftdlfekkgy tfdyaydwvg rpiptpvgsv hvdsgasait reshthrdrp
 361 sqqqplrnqv vsstngelnv ddptgahsna pitahaevev veeakcccff krkrkktaqr
 421 hk
```

FIGURE 3D: Predicted nucleotide sequence encoding human casein kinase 1, gamma 2 (SEQ ID NO:3)

```
   1 gggatttgca cggcagcaga gtcaccgtgg agaggccagg gtatcacaaa cttatggatt
  61 ttgacaagaa aggagggaaa ggggagacgg aggagggccg gagaatgtcc aaggccggcg
 121 ggggccggag cagccacggc atccggagct cggggaccag ctcgggggtc ctgatggtgg
 181 gccccaactt ccgcgtcggc aagaagatcg gctgcggcaa cttcggggag ctccgcctag
 241 gaaagaatct ctatacaaat gaatacgtgg ctatcaaatt ggagccgatc aagtcccggg
```

```
 301 ccccgcagct gcacctggag taccggttct acaagcagct cagcgccaca gagggcgtcc
 361 ctcaggtcta ctacttcggt ccgtgcggga attacaacgc catggtgctg gagctgctgg
 421 ggcccagcct ggaggacctg ttcgacctgt gcgaccggac cttcacgctc aagacggtgc
 481 tgatgatcgc catccagctg atcacgcgca tggagtatgt gcacaccaag agcctaatct
 541 accgggacgt gaagcccgag aacttcctgg tgggccgccc ggggaccaag cggcagcatg
 601 ccatccacat catcgacttc gggctggcca aggagtacat cgaccccgag accaagaagc
 661 acatcccgta ccgcgagcac aagagcctga cgggcacggc gcgctacatg agcatcaaca
 721 cgcacctggg caaggagcag agccgccgcg acgacctgga ggcgctgggc cacatgttca
 781 tgtacttcct gcgcggcagc ctcccctggc aggggctcaa ggccgacacg ctcaaggagc
 841 ggtaccagaa gatcggggac accaaacgcg ccacgcccat cgaggtgctc tgcgagaact
 901 tcccagagga gatggccacg tacctgcgct atgtgcggcg cctggacttc ttcgagaagc
 961 ccgactatga ctacctgcgg aagctcttca ccgacctctt cgaccgcagt ggcttcgtgt
1021 tcgactatga gtacgactgg gccgggaagc ccctgccgac ccccatcggc accgtccaca
1081 ccgacctgcc ctcccagcct cagctccggg acaaaaccca gccgcacagc aaaaaccagg
1141 cgttgaactc caccaacggg gagctgaatg cggacgaccc cacggccggc cactccaacg
1201 ccccgatcac agcgcctgca gaggtggagg tggccgatga aaccaaatgc tgctgtttct
1261 tcaagaggag aaagagaaaa tcgctgcagc gacacaagtg accctgggcg cgtgcagccc
1321 cctgaatctt ctccgtgcag ccccttgggg cgcgaccttg tgcgaggccc tcggggccca
1381 cccacagcgg cccagggcca gaccctggct ggaagccaga acgcagactg caggggccgc
1441 gcctggctca ggcggcccca cccccgggac gtggggtcac ttccttcatg taagactttg
1501 gccgaaattt ctacacctgt gtctagtcct cccctccaag agcattaact atttaaaaca
1561 aggaaaagag gaaaaaaaaa acagaggccc gccctacccc actcctgccc ctccgtttct
1621 ttgctgaagt gagtagtgtg atcctggagg ccccccggcc tggccccgcc ccgccagccg
1681 cccccgttag cgtcataaag tccagcttgt ctccctcgat ccaaaggccg ttttctcgag
1741 gggagggcag gcccggcctg gaggggtgct gtggagctgt cttgcccagg ccctcctggg
1801 agggggacag gcattgttgc caggggtgag gccgtgcccc aggcctcccc gaaaccaaag
1861 gggaaggcag gggtggggcc gtggctgaag ccggctcccc aaccaaaatg ctgcaccaaa
1921 gctcgggcgc cgcgggcacg gctgctgcag tctcttccca gcctggccct ggcaaggggc
1981 gggtgggcgc tgccaggcgg gtgcttctcg acgcacttgc tcccggaggc tgcgccccgg
2041 cgcctggaac ccgaggtggg aggaccggtt ggtgtcaccc tgctcggccc tcagccctgc
2101 cgcgtggggc gcgtgggcac ggagcttcct gcctctgctc cgacacccgg caagcagccg
2161 gagacaaaac gccttaaagc ccccggccca gccctgcagg tatattgcag gggcctgggg
2221 gcggccctgg actggcgggc ggttccccag tggggtgccc tggaggctgc cgggcagagt
2281 ggagcagctt ggggccgtgc ccagggcggt ggctgtgagt ctagtttttg ctttaccaag
2341 tgtacagaaa tggcatttac gtttctctga tgctcccttg aagccataga atttaggggc
2401 ttttttaaaa aaataaaaga aaaatgaaac caaaaaaaaa aaaaaa
```

FIGURE 3E: Predicted amino acid sequence of human casein kinase 1, gamma 2 (SEQ ID NO:4)

```
  1 mdfdkkggkg eteegrrmsk agggrsshgi rssgtssgvl mvgpnfrvgk kigcgnfgel
 61 rlgknlytne yvaiklepik srapqlhley rfykqlsate gvpqvyyfgp cgnynamvle
121 llgpsledlf dlcdrtftlk tvlmiaiqli trmeyvhtks liyrdvkpen flvgrpgtkr
181 qhaihiidfg lakeyidpet kkhipyrehk sltgtaryms inthlgkeqs rrddlealgh
241 mfmyflrgsl pwqglkadtl keryqkigdt kratpievlc enfpeematy lryvrrldff
301 ekpdydylrk lftdlfdrsg fvfdyeydwa gkplptpigt vhtdlpsqpq lrdktqphsk
361 nqalnstnge lnaddptagh snapitapae vevadetkcc cffkrrkrks lqrhk
```

FIGURE 3F: Predicted nucleotide sequence encoding human casein kinase 1, gamma 3 (SEQ ID NO:5)

```
  1 gaattcaaag tggagtaccg caaacttgat atggaaaata aaaagaaaga caaggacaaa
```

```
  61 tcagatgata gaatggcacg acctagtggt cgatcgggac acaacactcg aggaactggg
 121 tcttcatcgt ctggagtttt aatggttgga cctaacttta gagttggaaa aaaaattgga
 181 tgtggcaatt ttggagaatt acgattaggg aaaaatttat acacaaatga atatgtggca
 241 attaagttgg agcccatgaa atcaagagca ccacagctac atttggaata cagattctat
 301 aagcagttag gatctggaga tggtatacct caagtttact atttcggccc ttgtggtaaa
 361 tacaatgcta tggtgctgga actgctggga cctagtttgg aagacttgtt tgacttgtgt
 421 gacagaacat tttctcttaa aacagttctc atgatagcta tacaactgat ttctcgcatg
 481 gaatatgtcc attcaaagaa cttgatatac agagatgtaa aacctgagaa cttcttaata
 541 ggacgaccaa gaaacaaaac ccagcaagtt attcacatta tagattttgg tttggcaaag
 601 gaatatattg atccggagac aaagaaacac ataccataca gagaacacaa gagccttaca
 661 ggaacagcta gatatatgag cataaacaca catttaggaa aagaacaaag tagaagagac
 721 gatttagaag ctttaggtca tatgttcatg tattttctga gaggcagtct tccttggcaa
 781 ggcttaaagg ctgacacatt aaaggagagg tatcagaaaa ttggagatac aaaacgggct
 841 acaccaatag aagtgttatg tgaaaatttt ccagaaatgg caacatatct tcgttatgta
 901 agaaggctag atttttttga aaaaccagac tatgaatact taagaaagct ttttactgac
 961 ttgtttgatc gaaaaggata tatgtttgat tatgaatatg actggattgg taaacagttg
1021 cctactccag tgggtgcagt tcagcaagat cctgctctgt catcaaacag agaagcacat
1081 caacacagag ataagatgca acaatccaaa aaccagtcgg cagaccacag ggcagcttgg
1141 gactcccagc aggcaaatcc ccaccatttg agagctcacc ttgcagcaga cagacatggt
1201 ggctcggtac aggttgtaag ttctacaaat ggagagttaa acacagatga ccccaccgca
1261 ggacgttcaa atgcacccat cacagcccct actgaagtag aagtgatgga tgaaaccaag
1321 tgctgctgct ttttcaaacg aaggaaaagg aaaaccatac agcgccacaa atgactctgg
1381 acacagacag atcctgggga gttacttaca tgttcatctg ctgtcttgtg attaaaatca
1441 tctctgtagt gaccacgtat attttcaagg actcactctt agaaacaaaa atgtcatact
1501 atcatacttc attttgtggt tgtcttacat tctttttctt ttttttttc tctaatttaa
1561 cctttatgga agctttaaag ttttgtcaaa acatgagtgc tttgcccatc agtgaatgga
1621 atggaccaat gaggtggtat caatgaatat agttccatag aacattttcc agaagttctt
1681 ctgttgtaga aagcagtaca gtatcttaag tgtcaaccag ttatatacct aatctggttt
1741 tttataactt ctgtaagagc ataatcaaac aggaattttc ttttctcagt ggataataca
1801 acagagaaaa cagagttgcc caaatattta aaagaagtta ttccttgaga agttcatatt
1861 ttgtgacatc tgcattgatt tcagtattac tgatggtact gttattcata agtcatatta
1921 acattctctc cgtgaaatca tggtacagtc actgcccaga ggtactgagg aaaagcaata
1981 tgggttcggc agatggtggt ggtaaaatga atcttaagga gtgtggtaaa tatgtgctcc
2041 gcttttgttg catcactatg tgaagtactg tgttgcagaa gtggcaaaag cgcttatttt
2101 taaaaatgca aaatatttgt acaatgtaac tttatgcttc caaataataa tgtatgttag
2161 acagcaagaa atgaatactt taaaaagtga tatatgttgg agttataaag aaatacacta
2221 aggagaggta gtaaatgtga accttgttgc agtgtataag gtggaagcct aaagaaatct
2281 caccgaaact tactgctgaa tgattacatt ctcccttaag cagaaaactt tggatgtgcc
2341 atgcaatggt gtctgtgtaa ttattttgct ctttgattaa aaaaaagacc cccagcaata
2401 aaaagtgggt cactctatgc c
```

FIGURE 3G: Predicted amino acid sequence of human casein kinase 1, gamma 3 (SEQ ID NO:6)

```
  1 menkkkdkdk sddrmarpsg rsghntrgtg ssssgvlmvg pnfrvgkkig cgnfgelrlg
 61 knlytneyva iklepmksra pqlhleyrfy kqlgsgdgip qvyyfgpcgk ynamvlellg
121 psledlfdlc drtfslktvl miaiqlisrm eyvhsknliy rdvkpenfli grprnktqqv
181 ihiidfglak eyidpetkkh ipyrehkslt gtarymsint hlgkeqsrrd dlealghmfm
241 yflrgslpwq glkadtlker yqkigdtkra tpievlcenf pematylryv rrldffekpd
301 yeylrklftd lfdrkgymfd yeydwigkql ptpvgavqqd palssnreah qhrdkmqqsk
361 nqsadhraaw dsqqanphhl rahlaadrhg gsvqvvsstn gelntddpta grsnapitap
421 tevevmdetk cccffkrrkr ktiqrhk
```

FIGURE 4. CLUSTAL W (1.83) Protein Sequence Alignment Analysis

```
CK1 g3 Hs MENKKK----DKDKSDDRMARPSGRSGHNTRGTGSSSS-GVLMVGPNFRVGKKIGCGNFG
CK1 g1 Hs MDHPSR----EKDERQRTTKPMAQRSAHCSRPSGSSSSSGVLMVGPNFRVGKKIGCGNFG
CK1 g2 Hs MDFDKKGGKGETEEGRRMSKAGGGRSSHGIRSSGTSSG--VLMVGPNFRVGKKIGCGNFG
CG6963 Dm ----------------------------MYSTRQSVSTTT-GVLMVGPNFRVGKKIGCGNFG

CK1 g3 Hs ELRLGKNLYTNEYVAIKLEPMKSRAPQLHLEYRFYKQLGS-GDGIPQVYYFGPCGKYNAM
CK1 g1 Hs ELRLGKNLYTNEYVAIKLEPIKSRAPQLHLEYRFYKQLGSAGEGLPQVYYFGPCGKYNAM
CK1 g2 Hs ELRLGKNLYTNEYVAIKLEPIKSRAPQLHLEYRFYKQLSA-TEGVPQVYYFGPCGNYNAM
CG6963 Dm ELRLGKNLYNNEHVAIKMEPMKSKAPQLHLEYRFYKLLGSHAEGVPEVYYFGPCGKYNAL

CK1 g3 Hs VLELLGPSLEDLFDLCDRTFSLKTVLMIAIQLISRMEYVHSKNLIYRDVKPENFLIGRPR
CK1 g1 Hs VLELLGPSLEDLFDLCDRTFTLKTVLMIAIQLLSRMEYVHSKNLIYRDVKPENFLIGRQG
CK1 g2 Hs VLELLGPSLEDLFDLCDRTFTLKTVLMIAIQLITRMEYVHTKSLIYRDVKPENFLVGRPG
CG6963 Dm VMELLGPSLEDLFDICGRRFTLKSVLLIAIQLLHRIEYVHSRHLIYRDVKPENFLIGRTS

CK1 g3 Hs NKTQQVIHIIDFGLAKEYIDPETKKHIPYREHKSLTGTARYMSINTHLGKEQSRRDDLEA
CK1 g1 Hs NKKEHVIHIIDFGLAKEYIDPETKKHIPYREHKSLTGTARYMSINTHLGKEQSRRDDLEA
CK1 g2 Hs TKRQHAIHIIDFGLAKEYIDPETKKHIPYREHKSLTGTARYMSINTHLGKEQSRRDDLEA
CG6963 Dm TKREKIIHIIDFGLAKEYIDLDTNRHIPYREHKSLTGTARYMSINTHMGREQSRRDDLEA

CK1 g3 Hs LGHMFMYFLRGSLPWQGLKADTLKERYQKIGDTKRATPIEVLCENFP-EMATYLRYVRRL
CK1 g1 Hs LGHMFMYFLRGSLPWQGLKADTLKERYQKIGDTKRNTPIEALCENFPEEMATYLRYVRRL
CK1 g2 Hs LGHMFMYFLRGSLPWQGLKADTLKERYQKIGDTKRATPIEVLCENFPEEMATYLRYVRRL
CG6963 Dm LGHMFMYFLRGSLPWQGLKADTLKERYQKIGDTKRATPIEVLCDGHPEEFATYLRYVRRL

CK1 g3 Hs DFFEKPDYEYLRKLFTDLFDRKGYMFDYEYDWIGKQLPTPVGAVQQDP-ALSSN-REAHQ
CK1 g1 Hs DFFEKPDYEYLRTLFTDLFEKKGYTFDYAYDWVGRPIPTPVGSVHVDSGASAIT-RESHT
CK1 g2 Hs DFFEKPDYDYLRKLFTDLFDRSGFVFDYEYDWAGKPLPTPIGTVHTDLPSQPQL-RDKTQ
CG6963 Dm DFFETPDYDFLRRLFQDLFDRKGYTDEGEFDWTGKTMSTPVGSLQTGHEVIISPNKDRHN

CK1 g3 Hs HRDKMQQSKNQSADHRAAWDSQQANPHHLRAHLAADRHGGSVQVVSSTNGELNTDDPTAG
CK1 g1 Hs HRDRPSQ--------------QQP----LRN-----------QVVSSTNGELNVDDPTGA
CK1 g2 Hs PHS----------------------------KN-----------QALNSTNGELNADDPTAG
CG6963 Dm VTAKTNAKGG-----VAAWPDVPKPGATLGNLTPADRHG-SVQVVSSTNGELNPDDPTAG

CK1 g3 Hs RSNAPITAPTEVEVMDETKCCCFFKRRKRKTIQRHK
CK1 g1 Hs HSNAPITAHAEVEVVEEAKCCCFFKRKRKKTAQRHK
CK1 g2 Hs HSNAPITAPAEVEVADETKCCCFFKRRKRKSLQRHK
CG6963 Dm HSNTPITQQPEVEVVDETKCCCFFKRKKKKSTRQK-
```

FIGURE 5. Expression of casein kinase 1, gamma in mammalian tissues

FIGURE 5A. Real-time PCR analysis of casein kinase 1, gamma 1 expression in wildtype mouse tissues (DCt pancreas = 20,57)

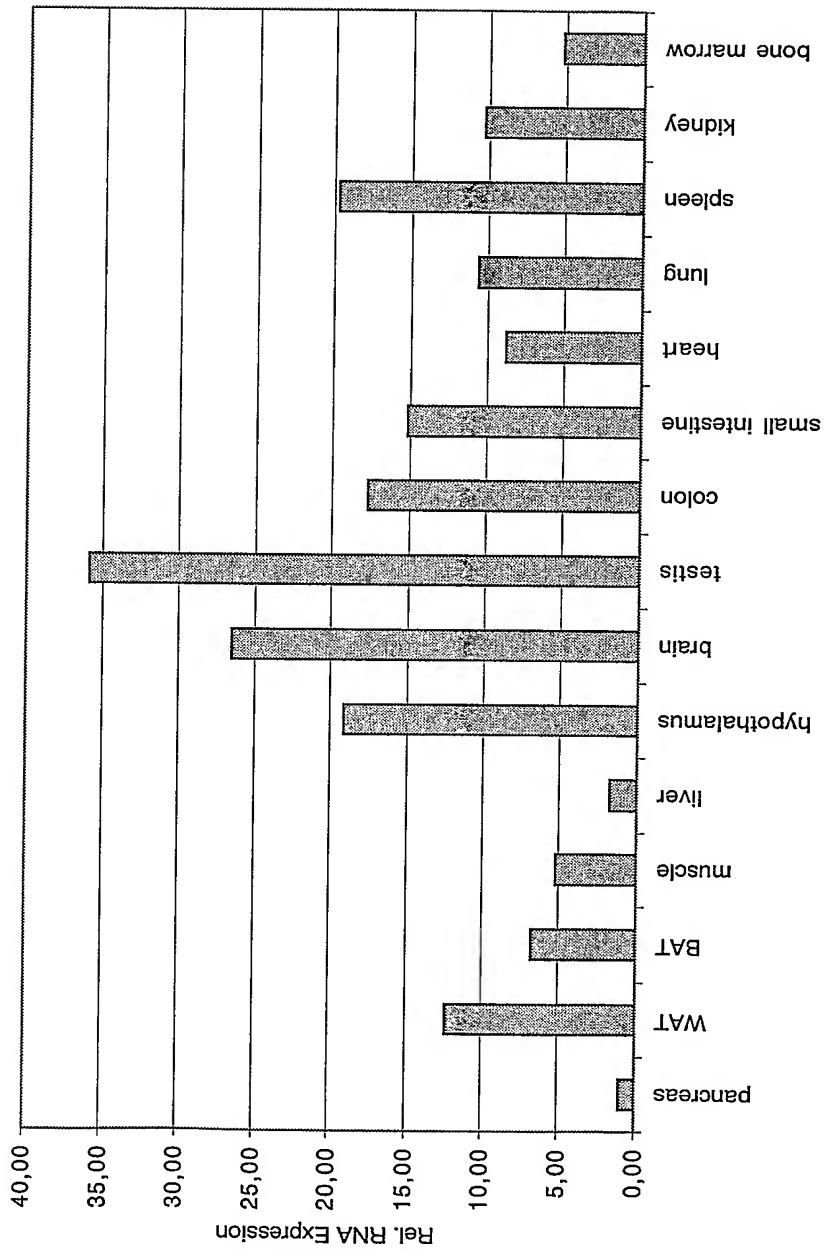
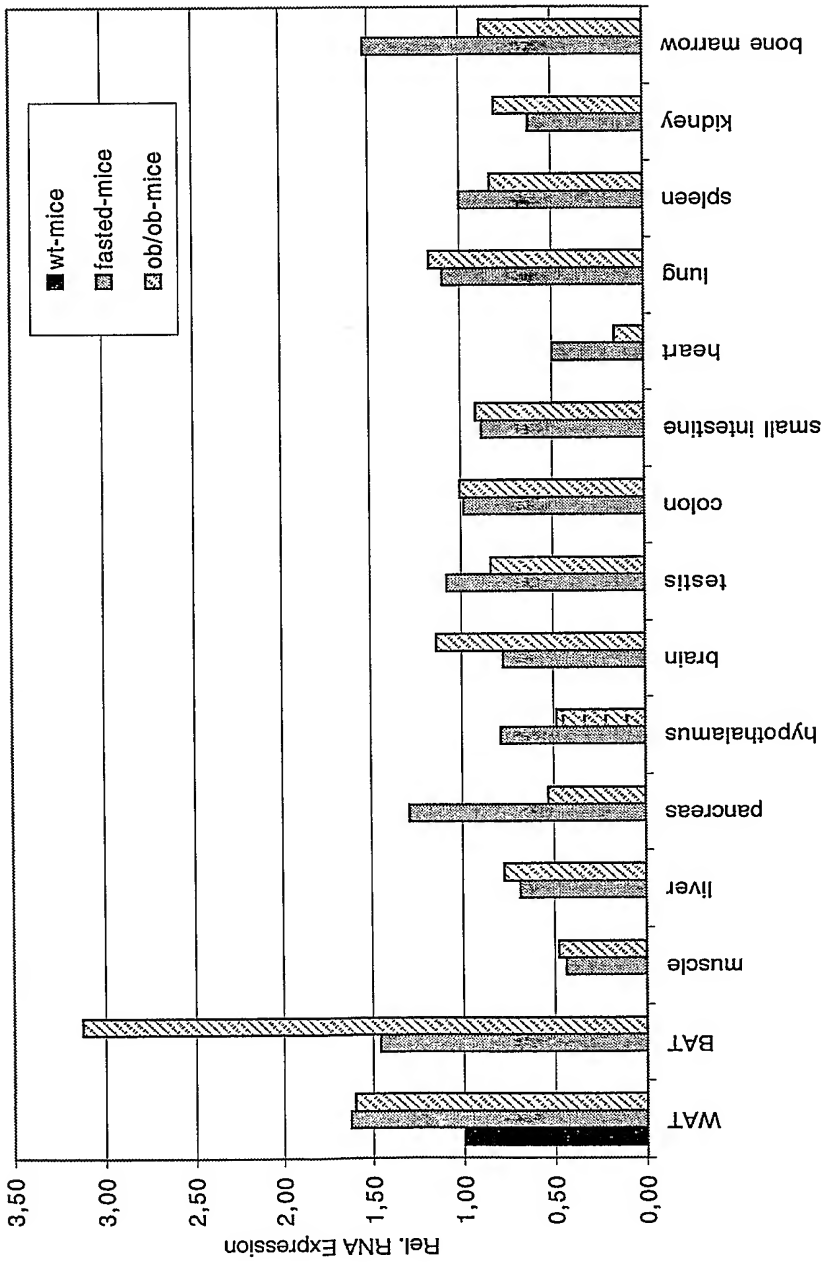


FIGURE 5B. Real-time PCR analysis of casein kinase 1, gamma 1 expression in different mouse models



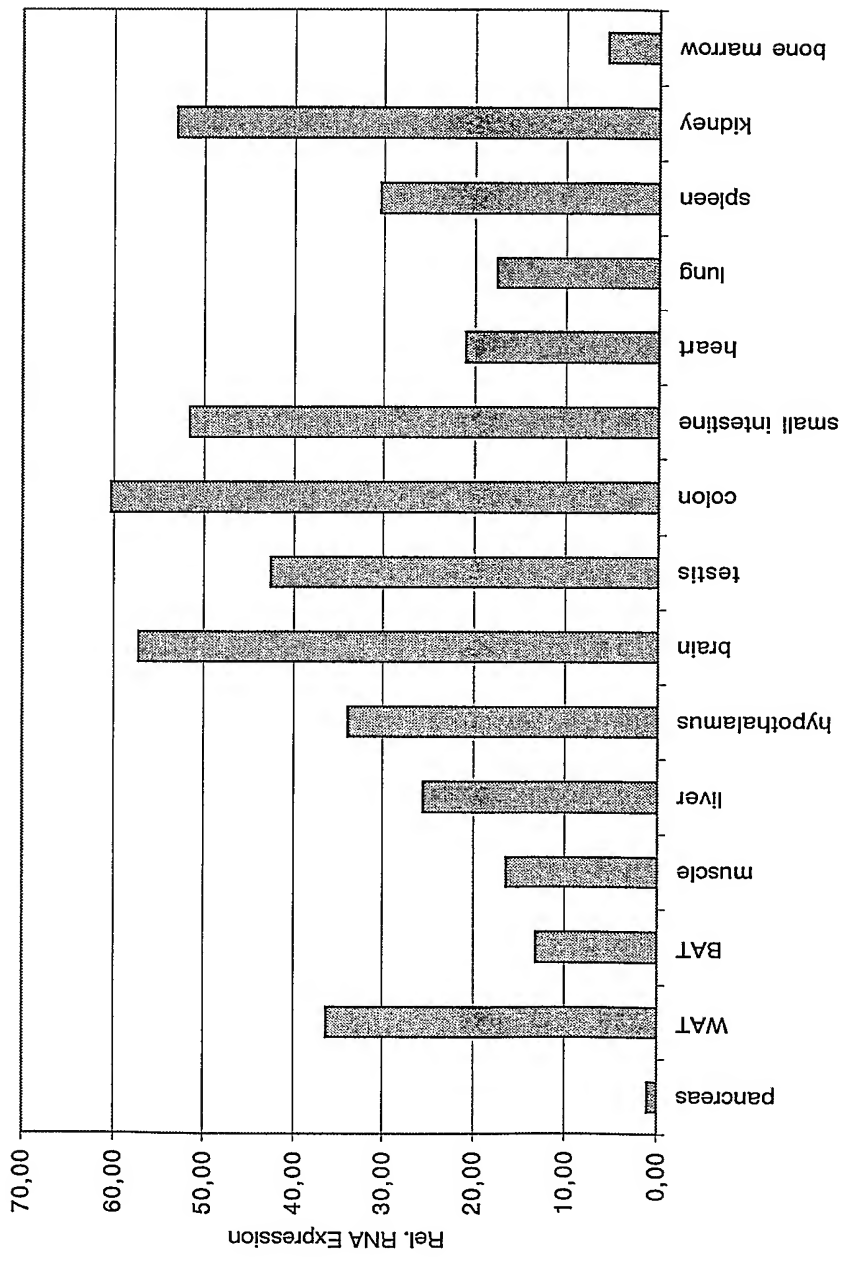


FIGURE 5C. Real-time PCR analysis of casein kinase 1, gamma 3 expression in wildtype mouse tissues (DCt pancreas = 17,92)

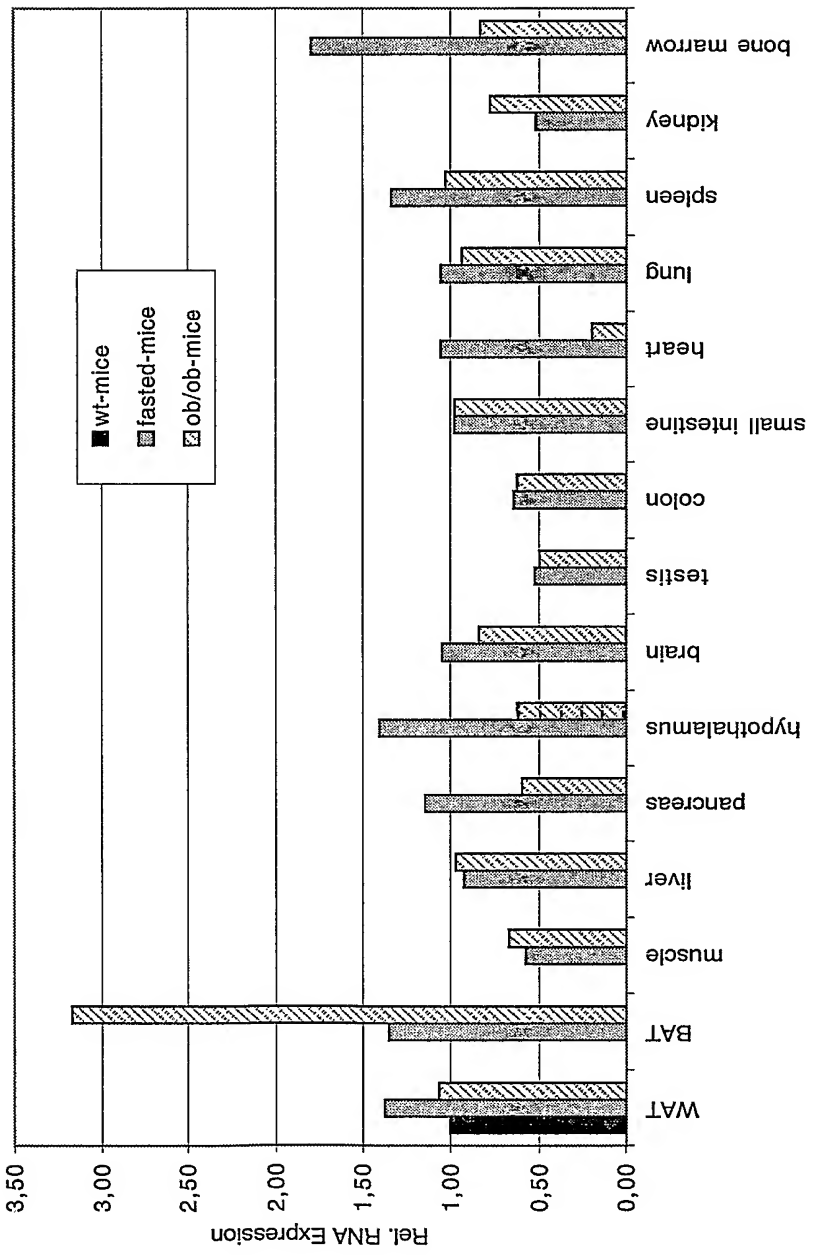
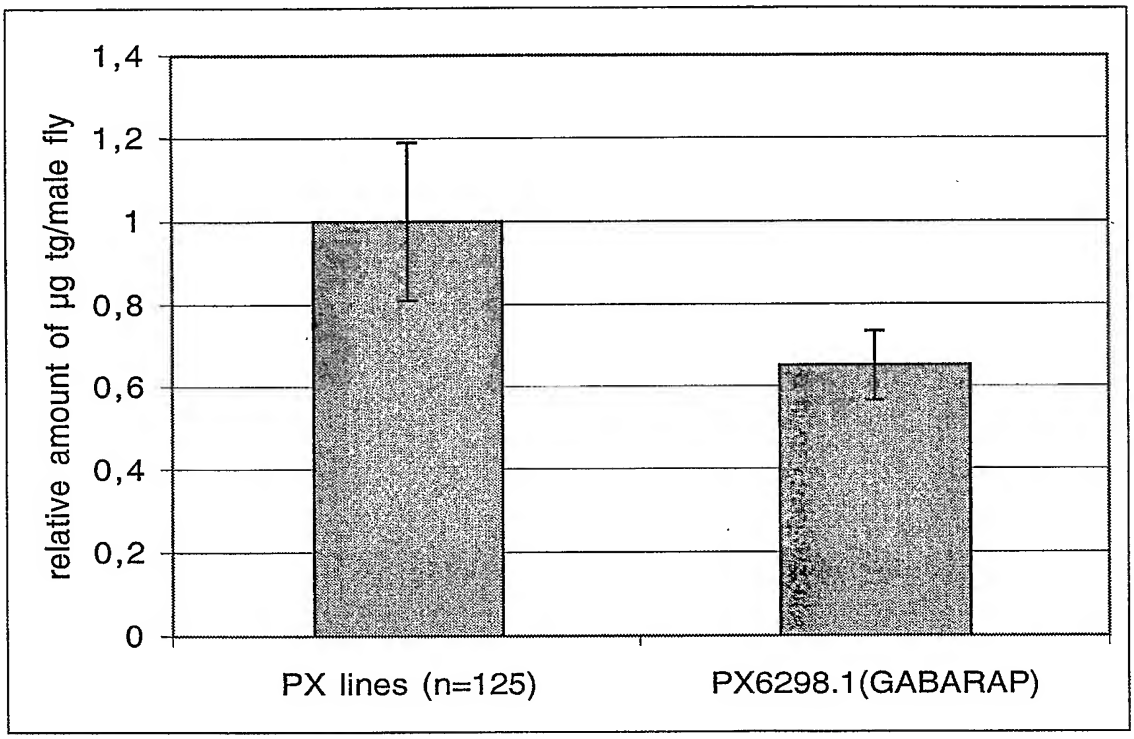


FIGURE 5D. Real-time PCR analysis of casein kinase 1, gamma 3 expression in different mouse models

FIGURE 6. Triglyceride levels of a CG1534 (Gadfly Accession Number) mutant



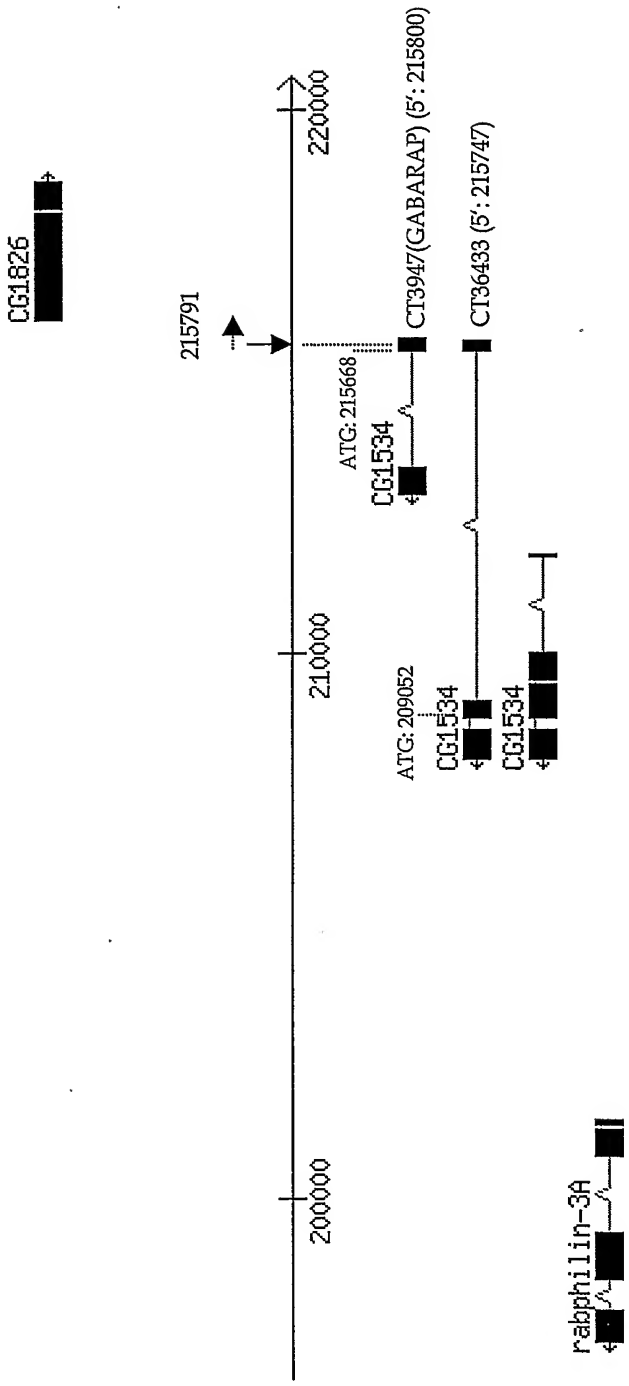


FIGURE 7. Molecular organisation of CG1534 (Gadfly Accession Number)

FIGURE 8: HUMAN HOMOLOG OF CG1534

FIGURE 8A. BLASTP search result for CG1534 (Gadfly Accession Number)

```
>gi|6005764|ref|NP_009209.1| GABA(A) receptor-associated protein [Homo
sapiens]
gi|9789961|ref|NP_062723.1| gamma-aminobutyric acid reseptor associated
protein; GABA-A receptor-associated protein [Mus musculus]

Query: Drosophila: CG1534 gene product, 121 amino acids; Human refers to
NP_009209.1;

Score =  200 bits (508), Expect = 2e-51; Identities = 107/117 (91%),
Positives = 113/117 (96%)

Droso: 1    MKFQYKEEHAFEKRRAEGDKIRRKYPDRVPVIVEKAPKARIGDLDKKKYLVPSDLTVGQF 60
            MKF YKEEH FEKRR+EG+KIR+KYPDRVPVIVEKAPKARIGDLDKKKYLVPSDLTVGQF
human: 1    MKFVYKEEHPFEKRRSEGEKIRKKYPDRVPVIVEKAPKARIGDLDKKKYLVPSDLTVGQF 60

Query: 61   YFLIRKRIHLRPEDALFFFVNNVIPPTSATMGSLYQEHHEEDYFLYIAYSDENVYGM 117
            YFLIRKRIHLR EDALFFFVNNVIPPTSATMG LYQEHHEED+FLYIAYSDE+VYG+
Sbjct: 61   YFLIRKRIHLRAEDALFFFVNNVIPPTSATMGQLYQEHHEEDFFLYIAYSDESVYGL 117
```

FIGURE 8B: Predicted nucleotide sequence encoding human GABA(A) receptor-associated protein (SEQ ID NO:7)

```
  1 gctccgctga atccgcccgc gcgtcgccgc cgtcgtcgcc gccccccgtc ccggcccccc
 61 tgggttccct cagcccagcc ctgtccagcc cggttcccgg gaggatgaag ttcgtgtaca
121 aagaagagca tccgttcgag aagcgccgct ctgagggcga gaagatccga aagaaatacc
181 cggaccgggt gccggtgata gtagaaaagg ctcccaaagc tcggatagga gacctggaca
241 aaaagaaata cctggtgcct tctgatctca cagttggtca gttctacttc ttgatccgga
301 agcgaattca tctccgagct gaggatgcct tgtttttctt tgtcaacaat gtcattccac
361 ccaccagtgc cacaatgggt cagctgtacc aggaacacca tgaagaagac ttctttctct
421 acattgccta cagtgacgaa agtgtctacg gtctgtgaag ctgctgcccc tgagctggag
481 gggggtctca ttctacaaag agagaggtgg ccccccttc ttgacctcct cctccttcaa
541 gctcaaacac cacctccctt attcaggacc ggcacttctt aatgtttgtg gctttctctc
601 cagcctctct taggaggggt aatggtggag ttggcatctt gtaactctcc tttctccttt
661 cttccccttt ctctgcccgc ctttcccatc ctgctgtaga cttcttgatt gtcagtctgt
721 gtcacatcca gtgattgttt tggtttctgt tccctttctg actgcccaag gggctcagaa
781 ccccagcaat cccttccttt cactaccttc ttttttgggg gtagttggaa gggactgaaa
841 ttgtgggggg aaggtaggag gcacatcaat aaagaggaaa ccaccaagct gaaaaaaaaa
901 aaaaaaaaaa aaaaaaaaaa aaaa
```

FIGURE 8C: Predicted amino acid sequence of human GABA(A) receptor-associated protein (SEQ ID NO:8)

```
 1 mkfvykeehp fekrrsegek irkkypdrvp vivekapkar igdldkkkyl vpsdltvgqf
61 yflirkrihl raedalfffv nnvipptsat mgqlyqehhe edfflyiays desvygl
```

FIGURE 8D: Predicted nucleotide sequence encoding human GABARAP like 1 (SEQ ID NO:9)

```
  1 cgtcacagcc cgacgcgcca cccagctgtt tttgtgctca caagctctag cgaaaagccg
 61 ccggtatttc tccatctggc tctcctctac ctccaggcag gctcacccga gatccccgcc
```

```
 121 ccgaaccccc cctgcacact cggcccagcg ctgttgcccc cggagcggac gtttctgcag
 181 ctattctgag cacaccttga cgtcggctga gggagcggga cagggtcagc ggcgaaggag
 241 gcaggccccg cgcggggatc tcggaagccc tgcggtgcat catgaagttc cagtacaagg
 301 aggaccatcc ctttgagtat cggaaaaagg aaggagaaaa gatccggaag aaatatccgg
 361 acagggtccc cgtgattgta gagaaggctc caaaagccag ggtgcctgat ctggacaaga
 421 ggaagtacct agtgccctct gaccttactg ttggccagtt ctacttctta atccggaaga
 481 gaatccacct gagacctgag gacgccttat tcttctttgt caacaacacc atccctccca
 541 ccagtgctac catgggccaa ctgtatgagg acaatcatga ggaagactat tttctgtatg
 601 tggcctacag tgatgagagt gtctatggga aatgagtggt tggaagccca gcagatggga
 661 gcacctggac ttgggggtag gggaggggtg tgtgtgcgcg acatggggaa agagggtggc
 721 tcccaccgca aggagacaga aggtgaagac atctagaaac attacaccac acacaccgtc
 781 atcacatttt cacatgctca attgatattt tttgctgctt cctcggccca gggagaaagc
 841 atgtcaggac agagctgttg gattggcttt gatagaggaa tggggatgat gtaagtttac
 901 agtattcctg gggtttaatt gttgtgcagt ttcatagatg ggtcaggagg tggacaagtt
 961 ggggccagag atgatggcag tccagcagca actccctgtg ctcccttctc tttgggcaga
1021 gattctattt ttgacatttg cacaagacag gtagggaaag gggacttgtg gtagtggacc
1081 atacctgggg accaaaagag acccactgta attgatgcat tgtggcccct gatcttccct
1141 gtctcacact tcttttctcc catcccggtt gcaatctcac tcagacatca cagtaccacc
1201 ccaggggtgg cagtagacaa caacccagaa atttagacag ggatctctta cctttggaaa
1261 ataggggtta ggcatgaagg tggttgtgat taagaagatg gttttgttat taaatagcat
1321 taaactggaa ttgacaagag tgttgagcat ccctgtctaa cctgctcttt ctctttggtg
1381 ccccttatct cacccccttcc ttggaattta ataagtctca ggcatttcca attgtagact
1441 aaaaccactc ttagcatctc ctctagtatt ttccatgtat caggacagag gtgtcttatg
1501 tagggagggg gcaagtatga agtaaggtaa ttatatacta ctctcattca ggattcttgc
1561 tcccatgctg ctgtcccttc aggctcacat gcacaggaat gctacatgat ggccagctgc
1621 ttccctcctt ggttatcatc cactgcagct gctagttaga aaggtttgga gggatgactt
1681 ttagtaaatc atggggattt tattgattta ttttcacttt tgggattttg tggggtggga
1741 gtggggagca ggaattgcac tcagacatga catttcaatt catctctgct aatgaaaagg
1801 gttctttctc ttgggggaaa tgtgtgtgtc agttctgtca gctgcaagtt cttgtataat
1861 gaagtcaatg ccatcaggcc aaggaaataa aataattgct taccttaaaa aaaaaaaaaa
1921 aaaaaaaaaa aaa
```

FIGURE 8E: Predicted amino acid sequence of human GABARAP like 1 (SEQ ID NO:10)

```
 1 mkfqykedhp feyrkkegek irkkypdrvp vivekapkar vpdldkrkyl vpsdltvgqf
61 yflirkrihl rpedalfffv nntipptsat mgqlyednhe edyflyvays desvygk
```

FIGURE 8F: Predicted nucleotide sequence encoding human GABARAP like 2 (SEQ ID NO:11)

```
   1 cgacagccgg aagtcccgcc tgccgtgtag tcgccgccgt cgctgccgct gccgctgccg
  61 ccgtcgttgt tgttgtgctc ggtgcgctga gctccgcggc tccgcgagcc ggttccgtcc
 121 ccttcccgcc gccgccatga agtggatgtt caaggaggac cactcgctgg aacacagatg
 181 cgtggagtcc gcgaagattc gagcgaaata tcccgacagg gttccggtga ttgtggaaaa
 241 ggtctcaggc tctcagattg ttgacattga caaacggaag tacttggttc catctgatat
 301 cactgtggct cagttcatgt ggatcatcag gaaaaggatc cagcttcctt ctgaaaaggc
 361 gatcttcctg tttgtggata agacagtccc acagtccagc ctaactatgg gacagcttta
 421 cgagaaggaa aaagatgaag atggattctt atatgtggcc tacagcggag agaacacttt
 481 tggcttctga gggccattgc tgggctaggt gcaccgtaac tgcttgtgta tcttgtaaat
 541 agccagccat tttcagttat tataccagaa cctcttcaca tagacctatt agtgcatttg
 601 taactggatt tatttcttaa tatattggaa ggttttgttt ccttagacta gtaaattatc
 661 atacagagtt ttattttgag tttttctttt tgtgcattgt cctcatgcct gtattctcca
 721 ggaaacttgt ccttctggaa atcatattga atgatatttc tatatcgaag tgaggtaggt
 781 gcggtattaa agtgaaaggg aaggtgatgc atttattctg ggttatgctt gaagtgttag
```

```
 841 atggctaagt attaaaatta tccaaattaa atccttagca gtcagaacac ttgcttcact
 901 agaatatgcc aactgccaat catgttggac tgagctaatt tgttcctctt tctgaaacta
 961 ttaaggtaaa taattaacaa taaaaattct cttataaagg caaaaaaaaa aaaaaaaaaa
1021 aaaaaaaaaa a
```

FIGURE 8G: Predicted amino acid sequence of human GABARAP like 2 (SEQ ID NO:12)

```
 1 mkwmfkedhs lehrcvesak irakypdrvp vivekvsgsq ivdidkrkyl vpsditvaqf
61 mwiirkriql psekaiflfv dktvpqsslt mgqlyekekd edgflyvays gentfgf
```

FIGURE 8H: Predicted nucleotide sequence encoding human GABARAP like 3 (SEQ ID NO:13)

```
   1 gaaaagccgc cggtatttct ccacctggct ctcctctacc tccaggcagg cgcacccgag
  61 gtccccctcc caccccacct tctgccctcc cgcacacttg gaccagtgct gttgacccgg
 121 aagcggacat ttctgcagct attctaagca cacgtcggcg gagggagcgg gacgtggcca
 181 gcggtcagcg gcgaaggagg caggccctgc gcggggatca cggaagccct gtgattcacc
 241 atgaagttcc agtacaagga ggtccatccc tttgagtatc ggaaaaagga aggagaaaag
 301 atccggaaga aatatccgga cagggtcccc ttgattgtag agaaggctcc aaaagcaagg
 361 gtgcctgatc tggacaggag gaagtaccta gtgccctccg accttaccga tggccagttc
 421 taccttttaa tccggaagag aatccacctg agacctgagg acgccttatt cttctttgtc
 481 aacaacacta tccctcccac tagtgctacc atgggccaac tatatgagga cagtcatgag
 541 gaagatgatt ttctgtatgt ggcctacagt aatgagagtg tctatgggaa atgagtggtt
 601 ggaagcccag cagatggaag cacctggact taggggtagg ggaggggtgt gtgtgtgact
 661 tggggaaaga gagggcggct cccaccgtga ggagacagaa ggtgaagaca tatagaaact
 721 ttacaccgca cacaccgtca acgcattttc acatgctcaa ctgatatttt ttgttgcttc
 781 cttggcccag ggagaaagca tgtcaggaca gagctgttgg attggctttg atagaggaat
 841 ggggatgatg taattttatg gcattcctga gatttaattt ttgtgcagtt tcatagaaag
 901 gtcggtcagg aggtggacaa gttggggtca gagatgatgg cagtccagca gcaactccct
 961 gtgctccctt ctctttgggc agagattctg tttttgacag ttgcacaaga caggtaggga
1021 aaggggactt gtggtagtgg gccatacctg gggacgaaaa gagacccact gtaattgatg
1081 catcgtggcc cccgatcttc cgtatcccac acttcttttc tcccatccca gttgcaatct
1141 cactcacaaa catcacagta ccaccccagg ggcggcagta gacaccaacc cagaaattta
1201 gacagggatc tcttatcttt ggaaaatagg ggttaggcat gagggtggtt atgattaaga
1261 agataatttt gttgttaaat agcattaaac tggaattgac agagtgagtt gagcatctct
1321 gtctaacctg ctctttctct ctggtgctcc tcatctcacc cctaccttgg aatttaataa
1381 gcttcaggca tttccaattg cagactaaaa ccacttctac catctcctct agtattttcc
1441 atgtatcagg acagagatgt cttatgtagg gaaggggcag gtatgaagtg aggtagatta
1501 tctatacctc tcactcattc aggattctcg ctcccatgct gctgtccctt cattctcaca
1561 ctcacaggaa tgctatgtga tggccagctg cttcccttct tggttatcca ctgcagctgc
1621 tagttagaaa ggtttgcagg gatgactttt agtaaatcat ggggatttta ttgatttatt
1681 atcacttata ggattttgtg gggtgggagt ggggagcagg aattgcactc agacatgaca
1741 tttcaattca tctctgcaaa tgaaaagggt tcttcctctt gggggaaatc tgtgtgtcag
1801 ttctgtcagc tgcaagttct tgtgtaatga agtcaatgct gtcaggccaa g
```

FIGURE 8I: Predicted amino acid sequence of human GABARAP like 3 (SEQ ID NO:14)

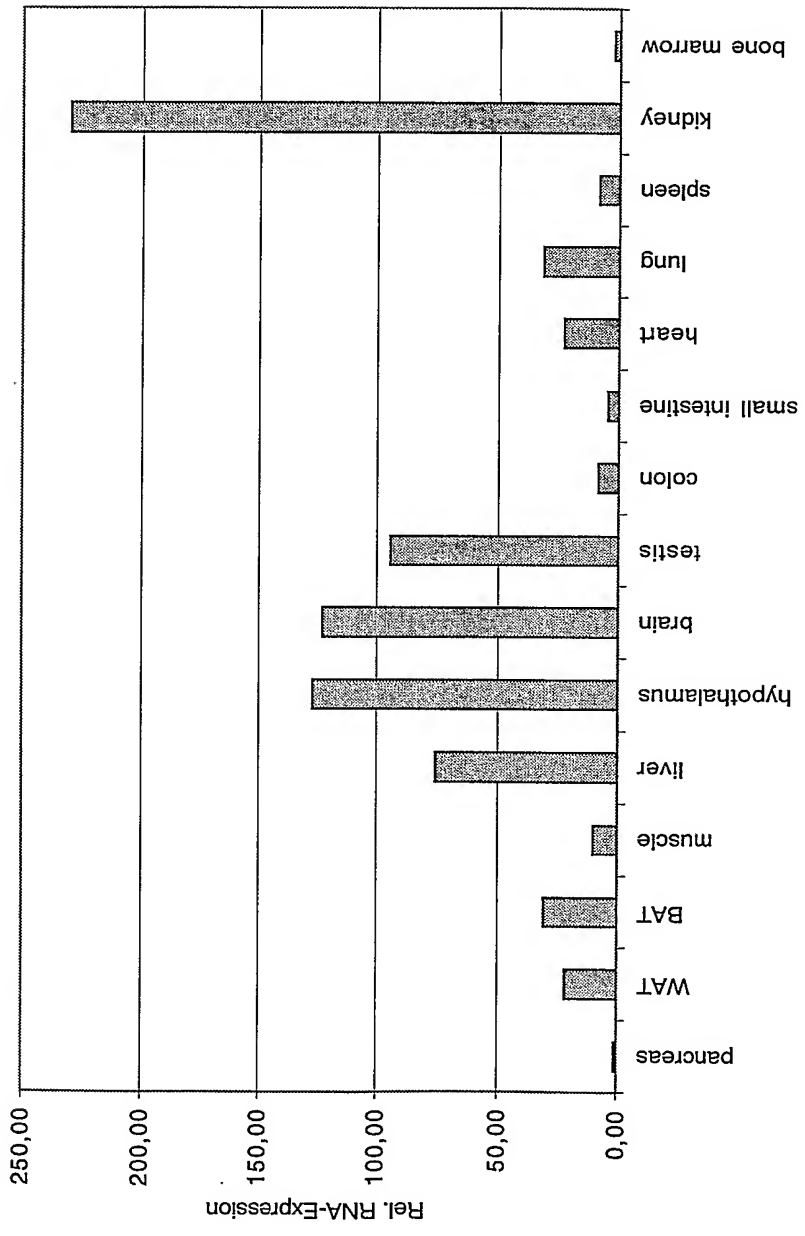
```
 1 mkfqykevhp feyrkkegek irkkypdrvp livekapkar vpdldrrkyl vpsdltdgqf
61 yllirkrihl rpedalfffv nntipptsat mgqlyedshe eddflyvays nesvygk
```

FIGURE 9. CLUSTAL W (1.82) Protein Sequence Alignment Analysis

```
GABARAP-l3 Hs MKFQYKEVHPFEYRKKEGEKIRKKYPDRVPLIVEKAPKARVPDLDRRKYLVPSDLTDGQF
GABARAP-l1 Hs MKFQYKEDHPFEYRKKEGEKIRKKYPDRVPVIVEKAPKARVPDLDKRKYLVPSDLTVGQF
GABARAP Hs    MKFVYKEEHPFEKRRSEGEKIRKKYPDRVPVIVEKAPKARIGDLDKKKYLVPSDLTVGQF
CG1534 Dm     MKFQYKEEHAFEKRRAEGDKIRRKYPDRVPVIVEKAPKARIGDLDKKKYLVPSDLTVGQF
CG12334 Dm    MNYQYKKDHSFDKRRNEGDKIRRKYPDRVPVIVEKAPKTRYAELDKKKYLVPADLTVGQF
GABARAP-l2 Hs MKWMFKEDHSLEHRCVESAKIRAKYPDRVPVIVEKVSGSQIVDIDKRKYLVPSDITVAQF

GABARAP-l3 Hs YLLIRKRIHLRPEDALFFFVNNTIPPTSATMGQLYEDSHEEDDFLYVAYSNESVYGK---
GABARAP-l1 Hs YFLIRKRIHLRPEDALFFFVNNTIPPTSATMGQLYEDNHEEDYFLYVAYSDESVYGK---
GABARAP Hs    YFLIRKRIHLRAEDALFFFVNNVIPPTSATMGQLYQEHHEEDFFLYIAYSDESVYGL---
CG1534 Dm     YFLIRKRIHLRPEDALFFFVNNVIPPTSATMGSLYQEHHEEDYFLYIAYSDENVYGMAKI
CG12334 Dm    YFLIRKRINLRPDDALFFFVNNVIPPTSATMGALYQEHFDKDYFLYISYTDENVYGRQ--
GABARAP-l2 Hs MWIIRKRIQLPSEKAIFLFVDKTVPQSSLTMGQLYEKEKDEDGFLYVAYSGENTFGF---

GABARAP-l3 Hs -
GABARAP-l1 Hs -
GABARAP Hs    -
CG1534 Dm     N
CG12334 Dm    -
GABARAP-l2 Hs -
```

FIGURE 10. Expression of GABARAP in mammalian tissues

FIGURE 10A. Real-time PCR analysis of GABARAP 2 expression in wildtype mouse tissues (DCt Pancreas = 11,83)



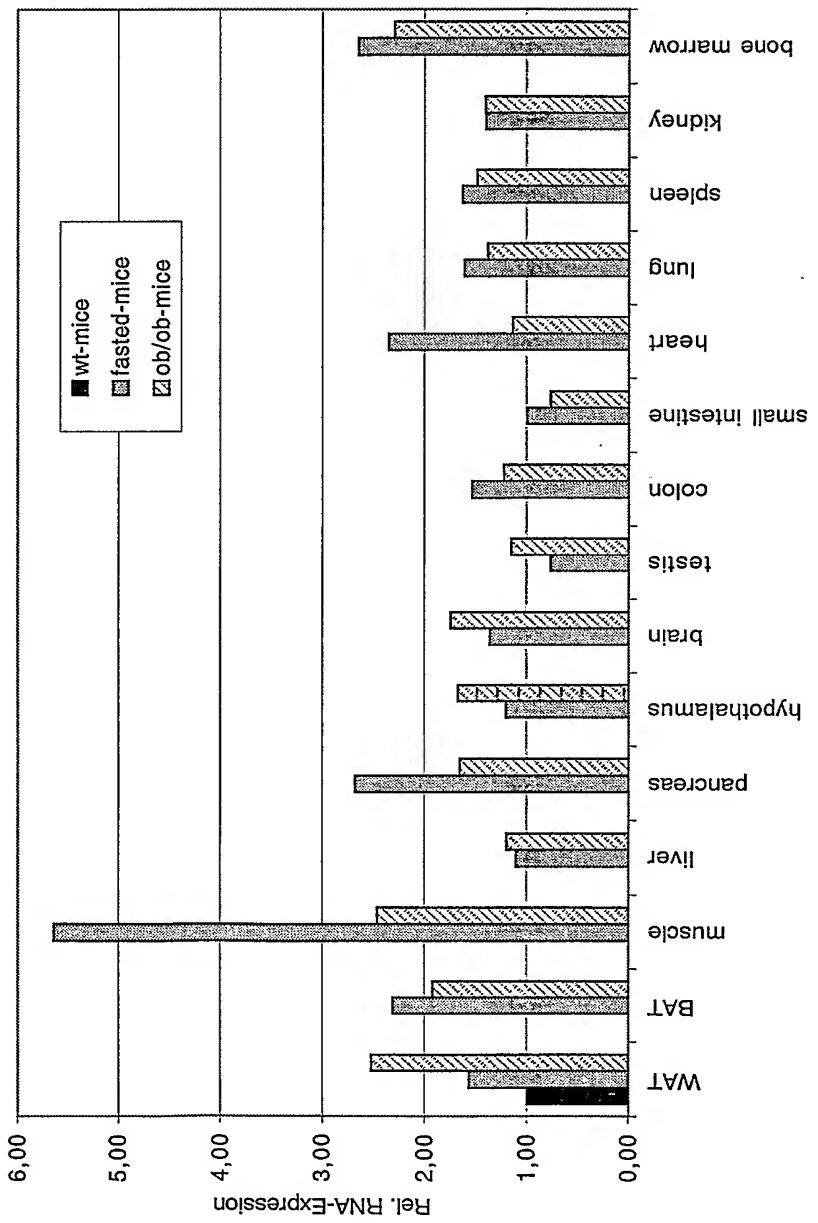


FIGURE 10B. Real-time PCR analysis of GABARAP 2 expression in different mouse models

FIGURE 10C. Real-time PCR analysis of GABARAP 2 expression in 3T3-F442A cells differentiated from preadipocytes to mature adipocytes (DCt(d0) = 7,11)

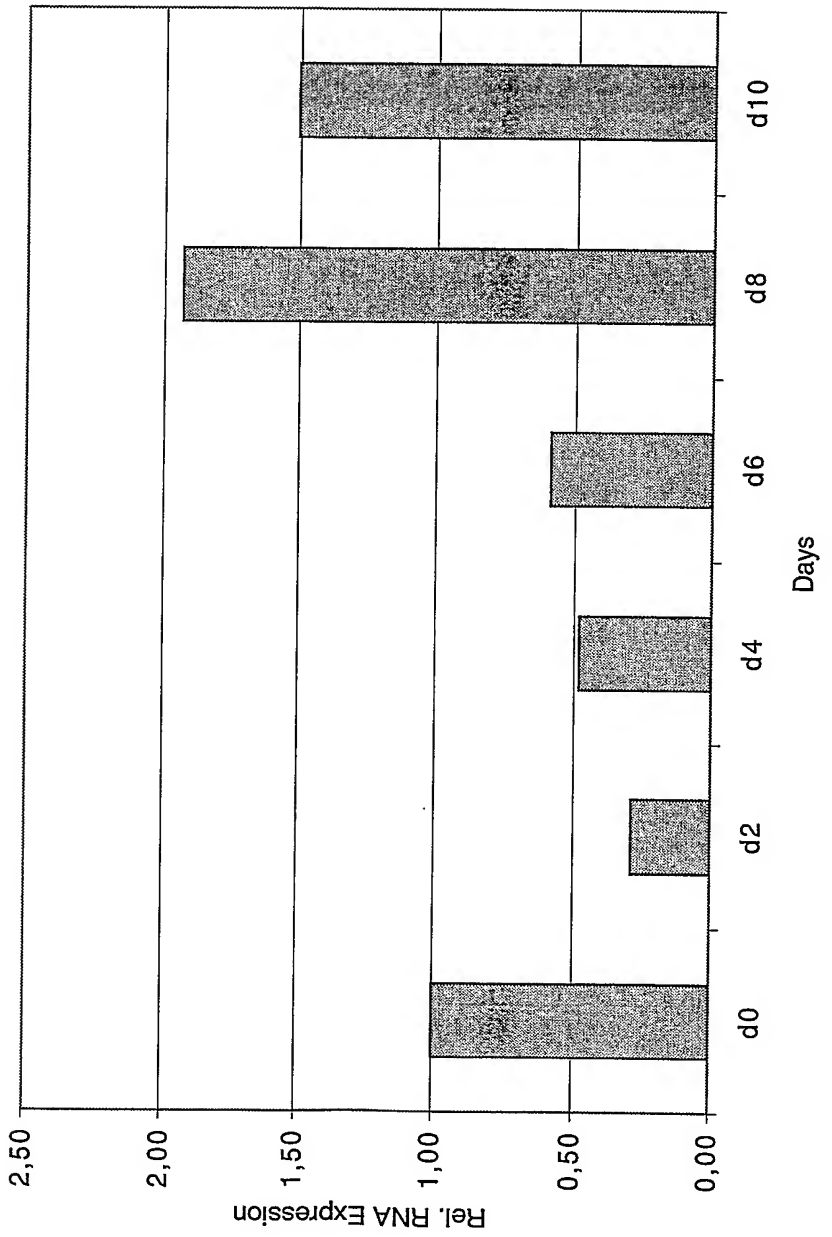


FIGURE 11. Triglyceride content of a CG10576 (Gadfly Accession Number) mutant

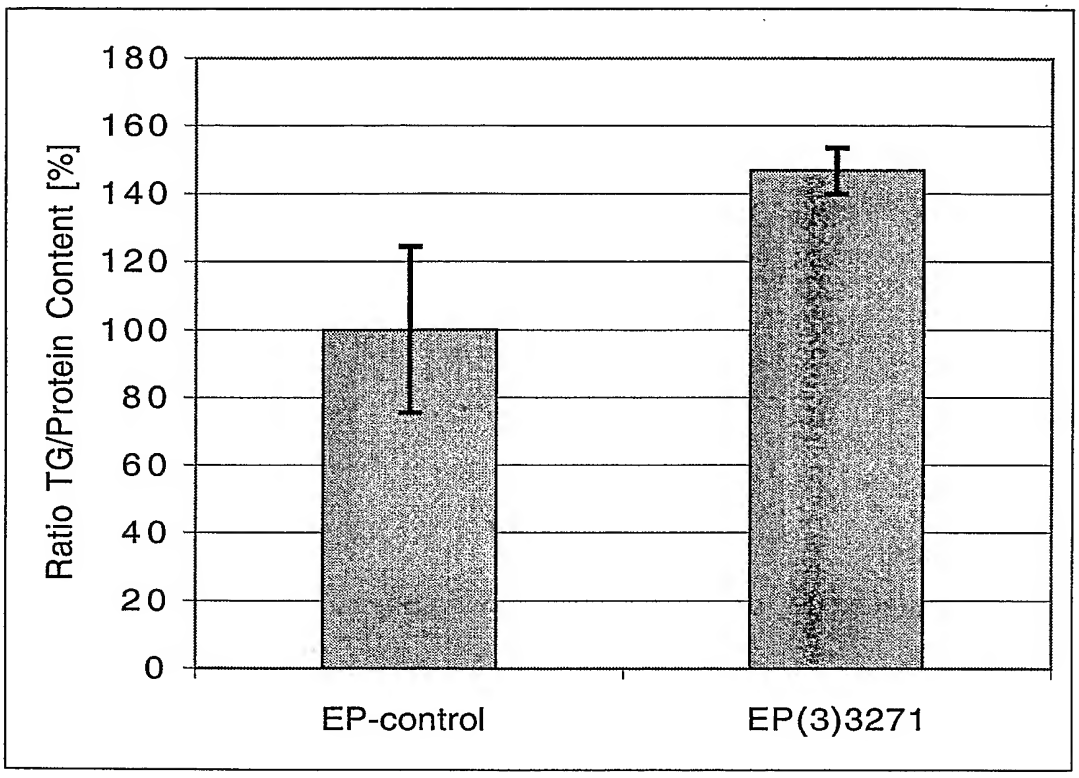


FIGURE 12. Molecular organisation of the gene with GadFly Accession Number CG10576

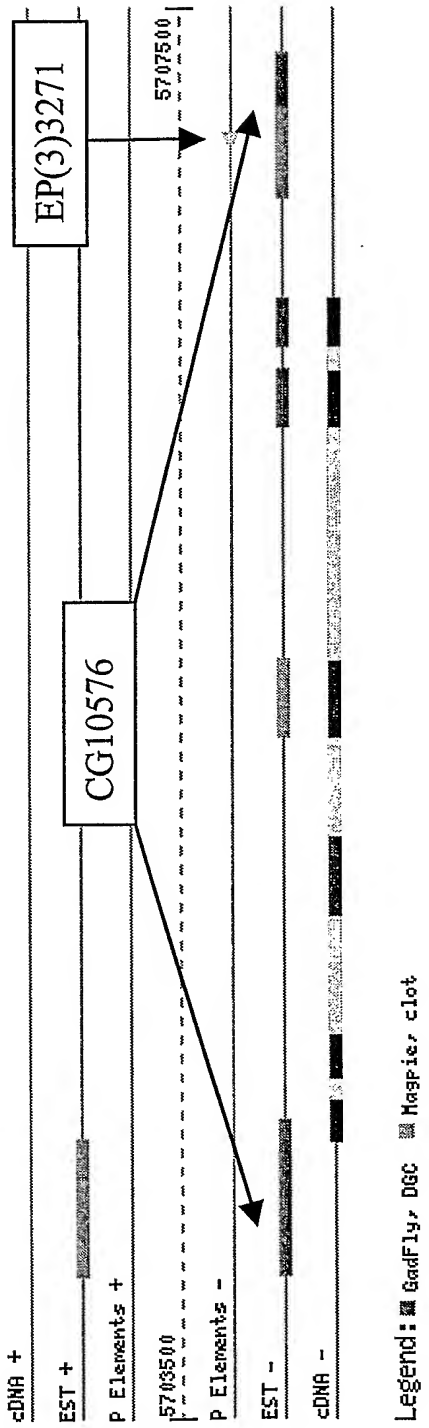


FIGURE 13: HUMAN HOMOLOG OF CG10576

FIGURE 13A. BLASTP result for CG10576 (Gadfly Accesssion Number)

Homology to human PA2G4 (gene ref XM_049048; protein ref XP_049048.1)

```
ref|XP_049048.1| (XM_049048) proliferation-associated 2G4, 38kD [Homo
sapiens]
sp|Q9UQ80|P2G4_HUMAN PROLIFERATION-ASSOCIATED PROTEIN 2G4 (CELL CYCLE
PROTEIN P38-2G4 HOMOLOG) (HG4-1)
gb|AAD05561.1| (AF104670) cell cycle protein [Homo sapiens]
gb|AAH01951.1|AAH01951 (BC001951) proliferation-associated 2G4, 38kD [Homo
sapiens] gb|AAH07561.1|AAH07561 (BC007561) Unknown (protein for MGC:15488)
[Homo sapiens], Length = 394

Score =  425 bits (1081), Expect = e-118
Identities = 212/386 (54%), Positives = 276/386 (70%), Gaps = 3/386 (0%)

Query: 1   MADVEKEPEKTIAEDLVVTKYKLAGEIVNKTLKAVIGLCVVDASVREICTQGDNQLTEET 60
           M+  +++ E+TIAEDLVVTKYK+ G+I N+ L++++        SV  +C +GD  + EET
Sbjct: 1   MSGEDEQQEQTIAEDLVVTKYKMGGDIANRVLRSLVEASSSGVSVLSLCEKGDAMIMEET 60

Query: 61  GKVYKKEKDLKKGIAFPTCLSVNNCVCHFSPAKNDADYTLKAGDVVKIDLGAHIDGFIAV 120
           GK++KKEK++KKGIAFPT +SVNNCVCHFSP K+D DY LK GD+VKIDLG H+DGFIA
Sbjct: 61  GKIFKKEKEMKKGIAFPTSISVNNCVCHFSPLKSDQDYILKEGDLVKIDLGVHVDGFIAN 120

Query: 121 AAHTIVVGAAADQKISGRQADVILAAYWAVQAALRLLKSGANNYSLTDAVQQISESYKCK 180
            AHT VV  A   +++GR+ADVI AA+   +AALRL+K G  N  +T+A  +++ S+ C
Sbjct: 121 VAHTFVVDVAQGTQVTGRKADVIKAAHLCAEAALRLVKPGNQNTQVTEAWNKVAHSFNCT 180

Query: 181 PIEGMLSHELKQFKIDGEKTIIQNPSEAQRKEHEKCTFETYEVYAIDVIVSTGEGVGREK 240
           PIEGMLSH+LKQ  IDGEKTIIQNP++ Q+K+HEK  FE +EVYA+DV+VS+GEG  ++
Sbjct: 181 PIEGMLSHQLKQHVIDGEKTIIQNPTDQQKKDHEKAEFEVHEVYAVDVLVSSGEGKAKDA 240

Query: 241 DTKVSIYKKS-EENYMLKMKASRALLAEVKTKYGNMPFNIRSFEEETKARMGVVECVGHK 299
             + +IYK+   + Y LKMK SRA  +EV+ ++  MPF +R+FE+E KARMGVVEC  H+
Sbjct: 241 GQRTTIYKRDPSKQYGLKMKTSRAFFSEVERRFDAMPFTLRAFEDEKKARMGVVECAKHE 300

Query: 300 MIEPFQVLYEKPSEIVAQFKHTVLLMPNGVNLVTGIPFEAENYVSEYSVAQEELKTLLAQ 359
           +++PF VLYEK  E VAQFK TVLLMPNG   +T  PFE + Y SE  V   ELK LL
Sbjct: 301 LLQPFNVLYEKEGEFVAQFKFTVLLMPNGPMRITSGPFEPDLYKSEMEVQDAELKALLQS 360

Query: 360 PLGPVKGKGKGKKA--TAGAATKVET 383
                K K KKA  TA  AT  ET
Sbjct: 361 SASRKTQKKKKKKASKTAENATSGET 386
```

FIGURE 13B: Predicted nucleotide sequence encoding human proliferation associated protein 2G4 (SEQ ID NO:15)

```
   1 ggatcgaggg gactctgacc acagcctgtg gctgggaagg gagacagagg cggcggcggc
  61 tcaggggaaa cgaggctgca gtggtggtag taggaagatg tcgggcgagg acgagcaaca
 121 ggagcaaact atcgctgagg acctggtcgt gaccaagtat aagatggggg gcgacatcgc
 181 caacagggta cttcggtcct tggtggaagc atctagctca ggtgtgtcgg tactcagcct
 241 gtgtgagaaa ggtgatgcca tgattatgga agaaacaggg aaaatcttca agaaagaaaa
 301 ggaaatgaag aaaggtattg cttttcccac cagcatttcg gtaaataact gtgtatgtca
 361 cttctccccc ttgaagagcg accaggatta tattctcaag gaaggtgact tggtaaaaat
 421 tgaccttggg gtccatgtgg atggcttcat cgctaatgta gctcacactt ttgtggttga
 481 tgtagctcag gggacccaag taacagggag gaaagcagat gttattaagg cagctcacct
 541 ttgtgctgaa gctgccctac gcctggtcaa acctggaaat cagaacacac aagtgacaga
 601 agcctggaac aaagttgccc actcatttaa ctgcacgcca atagaaggta tgctgtcaca
 661 ccagttgaag cagcatgtca tcgatggaga aaaaaccatt atccagaatc ccacagacca
 721 gcagaagaag gaccatgaaa aagctgaatt tgaggtacat gaagtatatg ctgtggatgt
 781 tctcgtcagc tcaggagagg gcaaggccaa ggatgcagga cagagaacca ctatttacaa
 841 acgagacccc tctaaacagt atggactgaa aatgaaaact tcacgtgcct tcttcagtga
 901 ggtggaaagg cgttttgatg ccatgccgtt tactttaaga gcatttgaag atgagaagaa
 961 ggctcggatg ggtgtggtgg agtgcgccaa acatgaactg ctgcaaccat ttaatgttct
1021 ctatgagaag gagggtgaat ttgttgccca gtttaaattt acagttctgc tcatgcccaa
1081 tggccccatg cggataacca gtggtccctt cgagcctgac ctctacaagt ctgagatgga
1141 ggtccaggat gcagagctaa aggccctcct ccagagttct gcaagtcgaa aaacccagaa
1201 aaagaaaaaa aagaaggcct ccaagactgc agagaatccc accagtgggg aaacattaga
1261 agaaaatgaa gctggggact gaggtgcgtc ccatctcccc agcttgctgc tcctgcctca
1321 tccccttccc accaaacccc agactctgtg aagtgcagtt cttctccacc taggaccgcc
1381 agcagagcgg ggggatctcc ctgcccccac cccagttccc caacccactc ccttccaaca
1441 acaaccagct ccaactgact ctggtcttgg gaggtgaggc ttcccaacca cggaagacta
1501 ctttaaacga aaaaaagaaa ttgaataata aaatcaggag tcaaaattca tcgtcttcaa
1561 ggcccctctt tctagccttt tctactactc tctgcttggt caaggtttgt gccccactac
1621 agaacagggc taaattagcc accaccactg aaaactcagc cgaatttttt tataccactc
1681 tgacgtcagc atttttt
```

FIGURE 13C: Predicted amino acid sequence of human human proliferation associated protein 2G4 (SEQ ID NO:16)

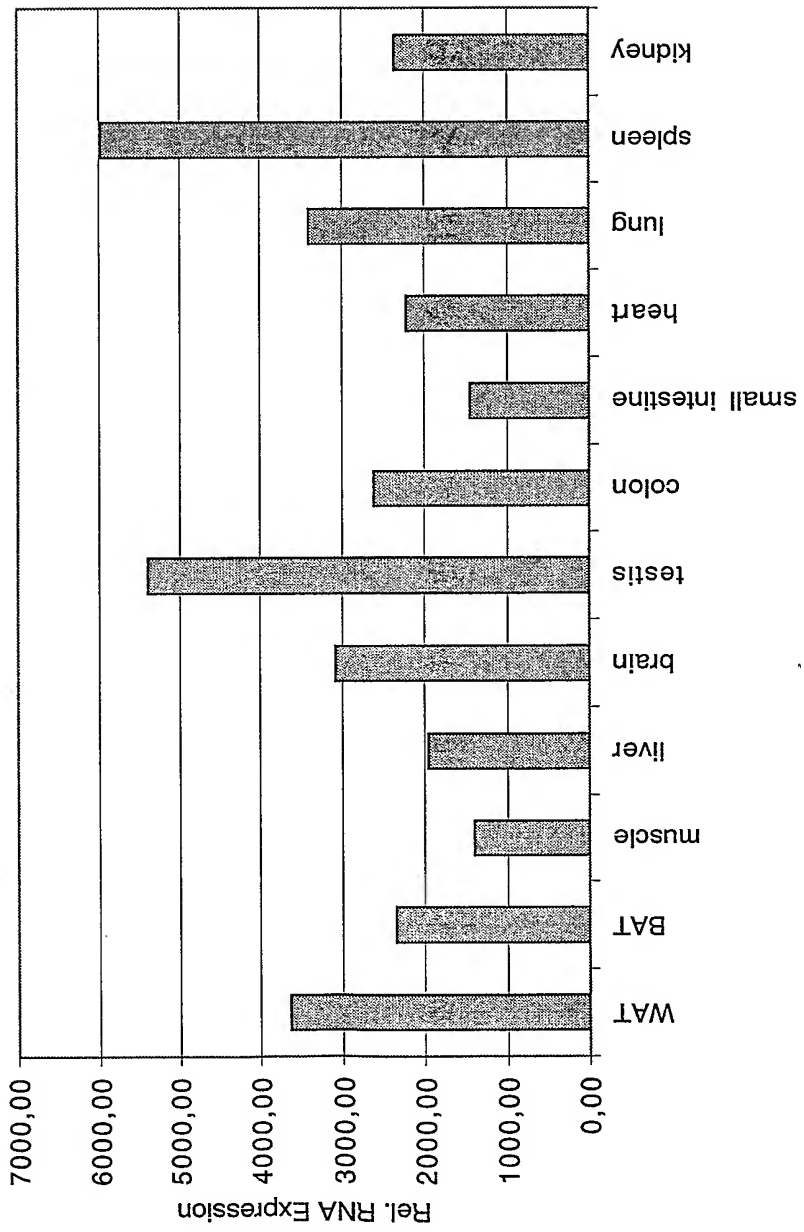
```
   1 msgedeqqeq tiaedlvvtk ykmggdianr vlrslveass sgvsvlslce kgdamimeet
  61 gkifkkekem kkgiafptsi svnncvchfs plksdqdyil kegdlvkidl gvhvdgfian
 121 vahtfvvdva qgtqvtgrka dvikaahlca eaalrlvkpg nqntqvteaw nkvahsfnct
 181 piegmlshql kqhvidgekt iiqnptdqqk kdhekaefev hevyavdvlv ssgegkakda
 241 gqrttiykrd pskqyglkmk tsraffseve rrfdampftl rafedekkar mgvvecakhe
 301 llqpfnvlye kegefvaqfk ftvllmpngp mritsgpfep dlyksemevq daelkallqs
 361 sasrktqkkk kkkasktaen ptsgetleen eagd
```

FIGURE 14. CLUSTAL W (1.7) Protein Sequence Alignment Analysis

```
CG10576         MADVEKEPEKTIAEDLVVTKYKLAGEIVNKTLKAVIGLCVVDASVREICTQGDNQLTEET
XP_049048.1     MSGEDEQQEQTIAEDLVVTKYKMGGDIANRVLRSLVEASSSGVSVLSLCEKGDAMIMEET
                *:. ::: *:************:.*:*.*:.*:::: .  ..** .:* :**  : ***

CG10576         GKVYKKEKDLKKGIAFPTCLSVNNCVCHFSPAKNDADYTLKAGDVVKIDLGAHIDGFIAV
XP_049048.1     GKIFKKEKEMKKGIAFPTSISVNNCVCHFSPLKSDQDYILKEGDLVKIDLGVHVDGFIAN
                **::****::********.:********** *.* ** ** **:******.*:*****

CG10576         AAHTIVVGAAADQKISGRQADVILAAYWAVQAALRLLKSGANNYSLTDAVQQISESYKCK
XP_049048.1     VAHTFVVDVAQGTQVTGRKADVIKAAHLCAEAALRLVKPGNQNTQVTEAWNKVAHSFNCT
                .***:**..* . :::**:**** **: ..:*****:*.* :* .:*:* :::::.*::*.

CG10576         PIEGMLSHELKQFKIDGEKTIIQNPSEAQRKEHEKCTFETYEVYAIDVIVSTGEGVGREK
XP_049048.1     PIEGMLSHQLKQHVIDGEKTIIQNPTDQQKKDHEKAEFEVHEVYAVDVLVSSGEGKAKDA
                ********:***. **********:: *:*:***. **.:****:**:**:*** .::

CG10576         DTKVSIYKKSE-ENYMLKMKASRALLAEVKTKYGNMPFNIRSFEEETKARMGVVECVGHK
XP_049048.1     GQRTTIYKRDPSKQYGLKMKTSRAFFSEVERRFDAMPFTLRAFEDEKKARMGVVECAKHE
                . :.:***:.  ::* ****:***:::**: ::. ***.:*:**:*.*********. *:

CG10576         MIEPFQVLYEKPSEIVAQFKHTVLLMPNGVNLVTGIPFEAENYVSEYSVAQEELKTLLAQ
XP_049048.1     LLQPFNVLYEKEGEFVAQFKFTVLLMPNGPMRITSGPFEPDLYKSEMEVQDAELKALLQS
                :::**:***** .*:*****.********   :*. ***.: * ** .* : ***:** .

CG10576         PLGPVKGKGKGKKATAGAATKVETAPAVETKA--
XP_049048.1     SASRKTQKKKKKKASKTAENATSGETLEENEAGD
                . .  . * * ***:  * . ..  .  *.:*
```

FIGURE 15. Expression of proliferation-associated 2G4 protein, 38kDa (PA2G4) in mammalian tissues

FIGURE 15A. Real-time PCR analysis of PA2G4 expression in wildtype mouse tissues (DCt ref36 = 25,62)



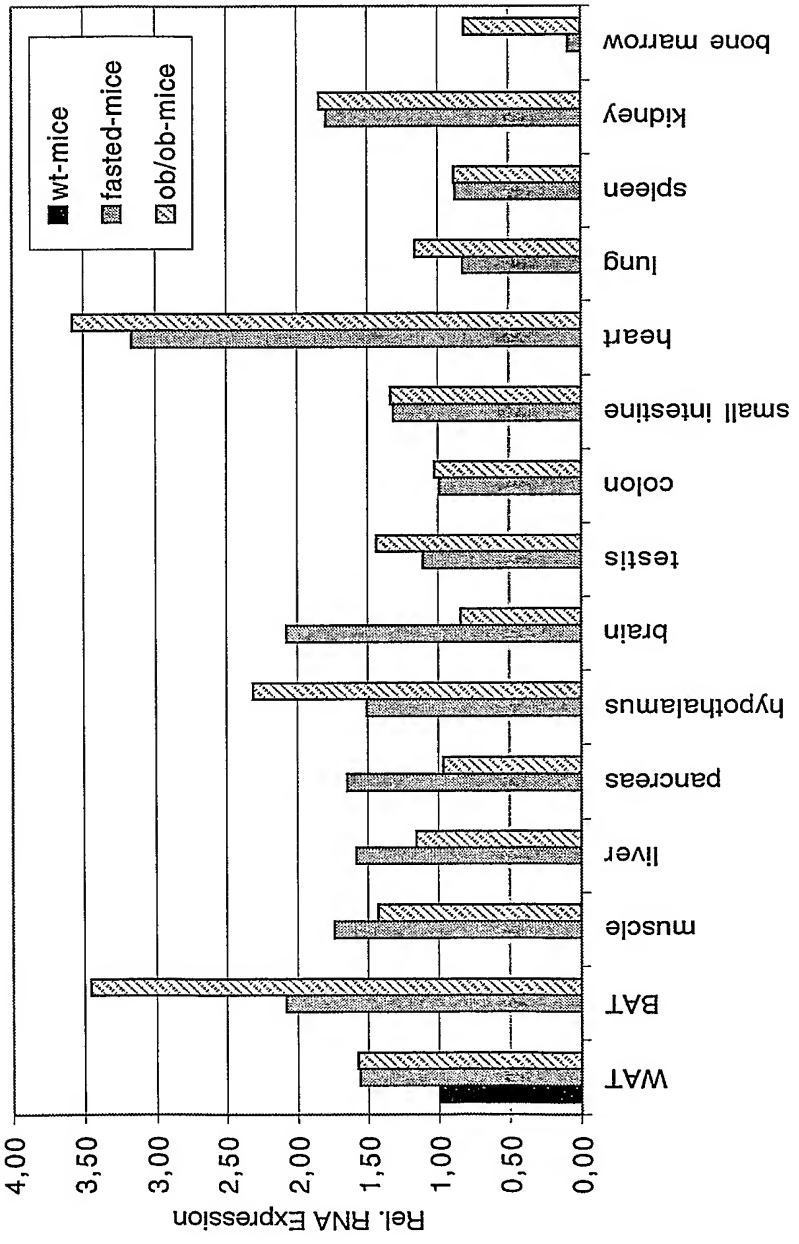


FIGURE 15B. Real-time PCR analysis of PA2G4 expression in different mouse models

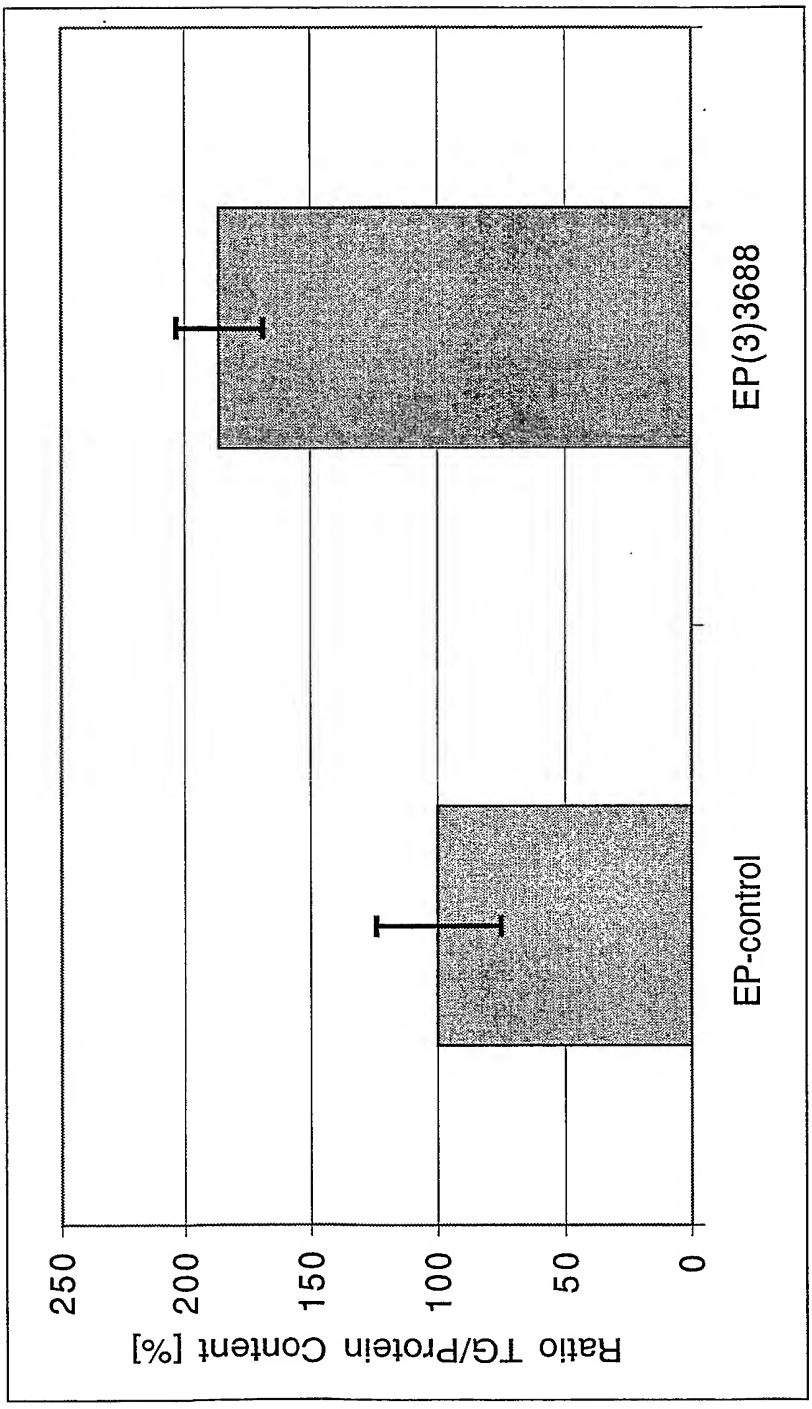


FIGURE 16. Triglyceride content of a *Mocs1* (Gadfly Accession Number CG7858) mutant

FIGURE 17. Molecular organisation of the *Mocs1* gene (Gadfly Accession Number CG7858)

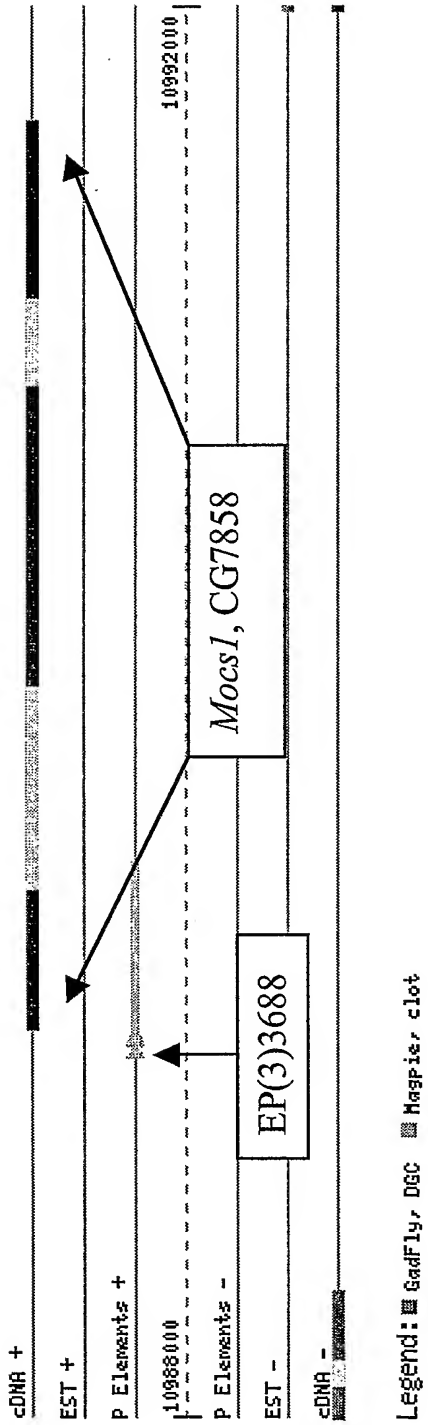


FIGURE 18: HUMAN HOMOLOG OF CG7858 (Mocs1)

FIGURE 18A. BLASTP search results for Mocs1 (Gadfly Accession Number CG7858)

```
gb|AAB87523.1| (AF034374) molybdenum cofactor biosynthesis protein A [Homo
sapiens]  Length = 385

Score =  445 bits (1132), Expect = e-123
Identities = 214/351 (60%), Positives = 274/351 (77%), Gaps = 7/351 (1%)

Query: 41  ATASVQPLEPEKQVLRKNSP-----LTDSFGRHHTYLRISLTERCNLRCDYCMPAEGVPL 95
           A A+ + +    +Q LR+++       LTDSFGR H+YLRISLTE+CNLRC YCMP EGVPL
Sbjct: 36  ARAASEEVSRRRQFLREHAAPFSAFLTDSFGRQHSYLRISLTEKCNLRCQYCMPEEGVPL 95

Query: 96  QPKNKLLTTEEILRLARIFVEQGVRKIRLTGGEPTVRRDIVEIVAQMKALPELEQIGITT 155
            PK  LLTTEEIL LAR+FV++G+ KIRLTGGEP +R D+V+IVAQ++ L  L  IG+TT
Sbjct: 96  TPKANLLTTEEILTLARLFVKEGIDKIRLTGGEPLIRPDVVDIVAQLQRLEGLRTIGVTT 155

Query: 156 NGLVLTRLLLPLQRAGLDNLNISLDTLKRDRFEKITRRKGWERVIAGIDLAVQLGYRP-K 214
           NG+ L RLL  LQ+AGL  +NISLDTL   +FE I RRKG+ +V+ GI  A++LGY P K
Sbjct: 156 NGINLARLLPQLQKAGLSAINISLDTLVPAKFEFIVRRKGFHKVMEGIHKAIELGYNPVK 215

Query: 215 VNCVLMRDFNEDEICDFVEFTRNRPVDVRFIEYMPFSGNKWHTERLISYKDTLQIIRQRW 274
           VNCV+MR  NEDE+ DF   T   P+DVRFIEYMPF GNKW+ ++++SYK+ L  +RQ+W
Sbjct: 216 VNCVVMRGLNEDELLDFAALTEGHPLDVRFIEYMPFDGNKWNFKKMVSYKEMLDTVRQQW 275

Query: 275 PDFKALPNGPNDTSKAYAVPGFKGQVGFITSMTEHFCGTCNRLRLTADGNIKVCLFGNKE 334
           P+ + +P   + T+KA+ +PGF+GQ+ FITSM+EHFCGTCNRLR+TADGN+KVCLFGN E
Sbjct: 276 PELEKVPEEESSTAKAFKIPGFQGQISFITSMSEHFCGTCNRLRITADGNLKVCLFGNSE 335

Query: 335 FSLRDAMRDESVSEEQLVDLIGAAVQRKKKQHAGMLNLSQMENRPMILIGG 385
            SLRD +R    SE++L+ +IGAAV RKK+QHAGM ++SQM+NRPMILIGG
Sbjct: 336 VSLRDHLR-AGASEQELLRIIGAAVGRKKRQHAGMFSISQMKNRPMILIGG 385


ref|XP_046687.1| (XM_046687) molybdenum cofactor synthesis 1 [Homo sapiens]
emb|CAA11897.1| (AJ224328) MOCS1A protein [Homo sapiens]
emb|CAC44527.1| (AJ293577) MOCS1A enzyme [Homo sapiens] Length = 385

Score =  444 bits (1129), Expect = e-123
Identities = 214/348 (61%), Positives = 272/348 (77%), Gaps = 7/348 (2%)

Query: 44  SVQPLEPEKQVLRKNSP-----LTDSFGRHHTYLRISLTERCNLRCDYCMPAEGVPLQPK 98
           S Q +   +Q LR+++       LTDSFGR H+YLRISLTE+CNLRC YCMP EGVPL PK
Sbjct: 39  SSQEVSRRRQFLREHAAPFSAFLTDSFGRQHSYLRISLTEKCNLRCQYCMPEEGVPLTPK 98

Query: 99  NKLLTTEEILRLARIFVEQGVRKIRLTGGEPTVRRDIVEIVAQMKALPELEQIGITTNGL 158
             LLTTEEIL LAR+FV++G+ KIRLTGGEP +R D+V+IVAQ++ L  L  IG+TTNG+
Sbjct: 99  ANLLTTEEILTLARLFVKEGIDKIRLTGGEPLIRPDVVDIVAQLQRLEGLRTIGVTTNGI 158

Query: 159 VLTRLLLPLQRAGLDNLNISLDTLKRDRFEKITRRKGWERVIAGIDLAVQLGYRP-KVNC 217
           L RLL  LQ+AGL  +NISLDTL   +FE I RRKG+ +V+ GI  A++LGY P KVNC
Sbjct: 159 NLARLLPQLQKAGLSAINISLDTLVPAKFEFIVRRKGFHKVMEGIHKAIELGYNPVKVNC 218

Query: 218 VLMRDFNEDEICDFVEFTRNRPVDVRFIEYMPFSGNKWHTERLISYKDTLQIIRQRWPDF 277
```

```
            V+MR  NEDE+ DF   T   P+DVRFIEYMPF GNKW+ ++++SYK+ L  +RQ+WP+
Sbjct: 219 VVMRGLNEDELLDFAALTEGLPLDVRFIEYMPFDGNKWNFKKMVSYKEMLDTVRQQWPEL 278

Query: 278 KALPNGPNDTSKAYAVPGFKGQVGFITSMTEHFCGTCNRLRLTADGNIKVCLFGNKEFSL 337
           + +P   + T+KA+ +PGF+GQ+ FITSM+EHFCGTCNRLR+TADGN+KVCLFGN E SL
Sbjct: 279 EKVPEEESSTAKAFKIPGFQGQISFITSMSEHFCGTCNRLRITADGNLKVCLFGNSEVSL 338

Query: 338 RDAMRDESVSEEQLVDLIGAAVQRKKKQHAGMLNLSQMENRPMILIGG 385
           RD +R    SE++L+ +IGAAV RKK+QHAGM ++SQM+NRPMILIGG
Sbjct: 339 RDHLR-AGASEQELLRIIGAAVGRKKRQHAGMFSISQMKNRPMILIGG 385
```

FIGURE 18B: Predicted nucleotide sequence encoding human molybdenum cofactor biosynthesis protein A and C (SEQ ID NO:17)

```
   1 cgctcgtatc aggcttcatg gcggcgcggc cactgtcccg gatgctgcgg cggcttctga
  61 ggtccagcgc ccggagctgc agctcagggg ctccggtgac ccagccctgc cccggggagt
 121 ccgcgcgagc tgcctcggag gaggtgtcca ggcggaggca gttcctgcgg gagcatgcgg
 181 ccccttctc cgccttcctc acagacagct tcggccggca gcacagctac ctgcggatct
 241 ccctcacaga gaagtgcaac ctcagatgtc agtactgcat gcccgaggag ggggtcccgc
 301 tgacccccaa agccaacctg ctgaccacag aggagatcct gaccctcgcc cggctctttg
 361 tgaaggaagg catcgacaag atccggctca caggtggaga gccgcttatc cggccggacg
 421 tggtggacat tgtggcccag ctccagcggc tggaagggct gagaaccata ggtgttacca
 481 ccaatggcat caacctggcc cggctactgc cccagcttca gaaggctggt ctcagtgcca
 541 tcaacatcag cctggacacc ctggtgcctg ccaagtttga gttcattgtc cgcaggaaag
 601 gcttccacaa ggtcatggag ggcatccaca aggccatcga gctgggctac aaccctgtga
 661 aggtgaactg tgtggtgatg cgaggcctta acgaggatga actcctggac tttgcggcct
 721 tgactgaggg ccacccccctg gatgtgcgct tcatagagta tatgcccttt gatggcaaca
 781 agtggaactt caagaagatg gtcagctata aggagatgct agacactgtc cggcagcagt
 841 ggccagagct ggagaaggtg ccagaggagg aatccagcac agccaaggcc tttaaaatcc
 901 ctggcttcca aggccagatc agcttcatca catccatgtc tgagcatttc tgtgggacct
 961 gcaaccgcct gcgaatcaca gctgatggga acctcaaggt ctgcctcttt ggaaactctg
1021 aggtatccct gcgggatcac ctgcgagctg gggcctctga gcaggagctg ctgagaatca
1081 ttggggctgc tgtgggcagg aagaagcggc agcatgcagg catgttcagt atttcccaga
1141 tgaagaaccg gcccatgatc ctcatcggtg ggtgacccat caagttattt ttgatgttcc
1201 ccaattcccc accagccaat ccaagcattt tctcctggga cccgctccat gttcagggtc
1261 taagacccag aatgagtttc tccagccagg tggccacttt atggaaagga tgcagggtcc
1321 cccagacccc tcctctagcc cagcagcggc tggggtctgg ctcctttcag agacactaca
1381 cttcccgtgc agactcagat gccaactcaa agtgccttag cccaggttcc tgggcttctg
1441 ctgccccctc aggaccccag ctaacctcag aacaactaac tcatgtggac tcggaaggac
1501 gggcagctat ggtagatgtg ggcaggaagc cagacacaga gcgggtggct gtggcttcag
1561 ccgtggtcct cctgggaccg gtagccttca agcttgtcca gcagaaccag ctcaagaaag
1621 gagatgccct agtggtggcc cagctggctg gagtccaggc agccaaggtg accagccagc
1681 tgatccctct gtgccaccac gtggccctga gccacatcca ggtgcagctg gagctggaca
1741 gcacacgcca tgccgtgaag atccaggcat cttgccgggc tcggggcccc accggggtgg
1801 agatggaggc cctgacctct gctgcagtgg ccgccctcac cctgtatgac atgtgcaagg
1861 ctgtcagcag ggacatcgtg ttggaggaga tcaagctcat tagcaagact ggtggtcagc
1921 gggggactt ccatcgggct tagcacctgc ccttctcacc catggcccac ccaggcctgg
1981 agctgggatg caatgtaggc tgagggaaag acgtcaggtt cctttaatca cagtcactgt
2041 ttgtttacct tgagcagtaa acccgaagtc agcctgctct actactaaca aacaggcctg
2101 ctgctagatg atctctaatg accaatgggg cttcctttct atagggagga taccagcagg
2161 cccttaagcc ttccaggaca ctaaggtcgt gggagcggga ctgcaacaag caatgccaga
2221 taactgagaa atcatgttct ttgtggacta tttcagacaa ccaggttccg acagtccagc
2281 ccagaacttt tccttctcat tttgggtttt ctcttctcct gctttcctgg ggagagatta
2341 agcgctcatt aagcagagga gcccactttg aggagagcaa agcacaagct tgcttgaaga
2401 atggatccca acttctcccc ggcagctctg cctccctaag tctgtgaagc cgcagccctg
2461 ccctgtcctg tcctgtcctg acttcatctc tccttctgcc caagtctgtg tcccatcaga
```

```
2521 cttgcagcct ttcagcttaa cagttgcccg gtcctgctgg ccccttttcc tctggccccc
2581 ctcttctgaa acaggatgtg cacacatggg ccatagccct aaggactcct gccagaccac
2641 acagcccaca cctggccctg ttcacggctg ttccacccac ccctctttat tctggagcat
2701 atcagggaaa gaaaagttga tgatagattg ccttcaccct cacagcgcac aaataaagct
2761 acgatgccaa ctttgcagat gcaagaatga agacactgtg tgggtagggc actgagctgc
2821 tgcagtttca cagggaaggc tgcacctatc aatcaatcaa tcaatcctat cccaagacac
2881 agttccctga gggaagaaga ggagggacct ggaaaggcct aagggtgtac tctctgtata
2941 gccccgctat gggaaaataa agtggagtag ggggcataga aaaaaaaaaa aaaaaaaaaa
3001 aaaaaaaaaa aaa
```

FIGURE 18C: Predicted amino acid sequence of human molybdenum cofactor biosynthesis protein A (SEQ ID NO:18)

```
  1 maarplsrml rrllrssars cssgapvtqp cpgesaraas eevsrrrqfl rehaapfsaf
 61 ltdsfgrqhs ylrisltekc nlrcqycmpe egvpltpkan lltteeiltl arlfvkegid
121 kirltggepl irpdvvdiva qlqrleglrt igvttnginl arllpqlqka glsainisld
181 tlvpakfefi vrrkgfhkvm egihkaielg ynpvkvncvv mrglnedell dfaalteghp
241 ldvrfieymp fdgnkwnfkk mvsykemldt vrqqwpelek vpeeesstak afkipgfqgq
301 isfitsmseh fcgtcnrlri tadgnlkvcl fgnsevslrd hlragaseqe llriigaavg
361 rkkrqhagmf sisqmknrpm iligg
```

FIGURE 18D: Predicted nucleotide sequence encoding human MOCS1 protein, isoform 1 (SEQ ID NO:19)

```
   1 gccagaaatc ttcccagtag agatcaccat ccgcccccga cccccaagct gaatacttaa
  61 ggggtgggtc cttcccatca agctgatttc tcaacgagag ggacaatccc agcttcccca
 121 acattgcaga gcccaaacat gtggaagagt tggaagctcc gcacagatgt cagagtaagg
 181 gagggggcag gcggttctcc ttgtgcctct tcccagcccg gtagcagggg cccatgcttc
 241 ctccctggtc tgtcctcgca ggaggtgtcc aggcggaggc agttcctgcg ggagcatgcg
 301 gcccccttct ccgccttcct cacagacagc ttcggccggc agcacagcta cctgcggatc
 361 tccctcacag agaagtgcaa cctcagatgt cagtactgca tgcccgagga gggggtcccg
 421 ctgaccccca aagccaacct gctgaccaca gaggagatcc tgaccctcgc ccggctcttt
 481 gtgaaggaag gcatcgacaa gatccggctc acaggtggag agccgcttat ccggccggac
 541 gtggtggaca ttgtggccca gctccagcgg ctggaagggc tgagaaccat aggtgttacc
 601 accaatggca tcaacctggc ccggctactg ccccagcttc agaaggctgg tctcagtgcc
 661 atcaacatca gcctggacac cctggtgcct gccaagtttg agttcattgt ccgcaggaaa
 721 ggcttccaca aggtcatgga gggcatccac aaggccatcg agctgggcta caaccctgtg
 781 aaggtgaact gtgtggtgat gcgaggcctt aacgaggatg aactcctgga ctttgcggcc
 841 ttgactgagg gcctcccccct ggatgtgcgc ttcatagagt atatgccctt tgatggcaac
 901 aagtggaact tcaagaagat ggtcagctat aaggagatgc tagacactgt ccggcagcag
 961 tggccagagc tggagaaggt gccagaggag gaatccagca cagccaaggc ctttaaaatc
1021 cctggcttcc aaggccagat cagcttcatc acatccatgt ctgagcattt ctgtgggacc
1081 tgcaaccgcc tgcgaatcac agctgatggg aacctcaagg tctgcctctt tggaaactct
1141 gaggtatccc tgcgggatca cctgcgagct ggggcctctg agcaggagct gctgagaatc
1201 attggggctg ctgtgggcag gaagaagcgg cagcatgcag gcatgttcag tatttcccag
1261 atgaagaacc ggcccatgat cctcatcggt gggtgaccca tcaagttatt tttgatgttc
1321 cccaattccc caccagccaa tccaagcatt ttctcctggg acccgctcca tgttcagggt
1381 ctaagaccca gaatgagttt ctccagccag gtggccactt tatggaaagg atgcagggtc
1441 ccccagaccc ctcctctagc ccagcagcgg ctggggtctg gctcctttca gagacactac
1501 acttcccgtg cagactcaga tgccaactca aagtgcctta gcccaggttc ctgggcttct
1561 gctgcccccct caggacccca gctaacctca gaacaactaa ctcatgtgga ctcggaagga
1621 cgggcagcta tggtagatgt gggcaggaag ccagacacag agcgggtggc tgtggcttca
1681 gccgtggtcc tcctgggacc ggtagccttc aagcttgtcc agcagaacca gctcaagaaa
1741 ggagatgccc tagtggtggc ccagctggct ggagtccagg cagccaaggt gaccagccag
1801 ctgatccctc tgtgccacca cgtggccctg agccacatcc aggtgcagct ggagctggac
```

```
1861 agcacacgcc atgccgtgaa gatccaggca tcttgccggg ctcggggccc caccggggtg
1921 gagatggagg ccctgacctc tgctgcagtg gccgccctca ccctgtatga catgtgcaag
1981 gctgtcagca gggacatcgt gttggaggag atcaagctca ttagcaagac tggtggtcag
2041 cgggggggact tccatcgggc ttagcacctg cccttctcac ccatggccca cccaggcctg
2101 gagctgggat gcaatgtagg ctgagggaaa gacgtcaggt tcctttaatc acagtcactg
2161 tttgtttacc ttgagcagta aacccgaagt cagcctgctc tactactaac aaacaggcct
2221 gctgctagat gatctctaat gaccaatggg gcttcctttc tatagggagg ataccagcag
2281 gcccttaagc cttccaggac actaaggtcg tgggagcggg actgcaacaa gcaatgccag
2341 ataactgaga aatcatgttc tttgtggact atttcagaca accaggttcc gacagtccag
2401 cccagaactt ttccttctca ttttgggttt tctcttctcc tgctttcctg gggagagatt
2461 aagcgctcat taagcagagg agcccacttt gaggagagca aagcacaagc ttgcctgaag
2521 aatggatccc aacttctccc cggcagctct gcctccctaa gtctgtgaag ccgcagccct
2581 gccctgtcct gtcctgtcct gacttcatct ctccttctgc ccaagtctgt gtcccatcag
2641 acttgcagcc tttcagctta acagttgccc ggtcctgctg gccccttttc ctctggcccc
2701 cctcttctga aacaggatgt gcacacatgg gccatagccc taaggactcc tgccagacca
2761 cacagcccac acctggccct gttcacggct gttccaccca cccctcttta ttctggagca
2821 tatcagggaa agaaaagttg atgatagatt gccttcaccc tcacagcgca caaataaagc
2881 tacgatgcca actttgc
```

FIGURE 18E: Predicted amino acid sequence of human MOCS1 protein, isoform 1 (SEQ ID NO:20)

```
  1 mwkswklrtd vrvregaggs pcassqpgsr gpcflpglss qevsrrrqfl rehaapfsaf
 61 ltdsfgrqhs ylrisltekc nlrcqycmpe egvpltpkan lltteeiltl arlfvkegid
121 kirltggepl irpdvvdiva qlqrleglrt igvttnginl arllpqlqka glsainisld
181 tlvpakfefi vrrkgfhkvm egihkaielg ynpvkvncvv mrglnedell dfaalteglp
241 ldvrfieymp fdgnkwnfkk mvsykemldt vrqqwpelek vpeeesstak afkipgfqgq
301 isfitsmseh fcgtcnrlri tadgnlkvcl fgnsevslrd hlragaseqe llriigaavg
361 rkkrqhagmf sisqmknrpm iligg
```

FIGURE 18F: Predicted nucleotide sequence of human MOCS1, isoform 2 protein (SEQ ID NO:21)

```
   1 gccagaaatc ttcccagtag agatcaccat ccgcccccga cccccaagaa tacttaaggg
  61 gtgggtcctt cccatcaagc tgatttctca acgagaggga caatcccagc ttccccaaca
 121 ttgcagagcc caaacatgtg gaagagttgg aagctccgca cagatgtcag agtaagggag
 181 ggggcaggcg gttctccttg tgcctcttcc cagcccggta gcaggggccc atgcttcctc
 241 cctggtctgt cctcgcagga ggtgtccagg cggaggcagt tcctgcggga gcatgcggcc
 301 cccttctccg ccttcctcac agacagcttc ggccggcagc acagctacct gcggatctcc
 361 ctcacagaga agtgcaacct cagatgtcag tactgcatgc ccgaggaggg ggtcccgctg
 421 acccccaaag ccaacctgct gaccacagag gagatcctga ccctcgcccg gctctttgtg
 481 aaggaaggca tcgacaagat ccggctcaca ggtggagagc cgcttatccg gccggacgtg
 541 gtggacattg tggcccagct ccagcggctg gaagggctga gaaccatagg tgttaccacc
 601 aatggcatca acctggcccg gctactgccc cagcttcaga aggctggtct cagtgccatc
 661 aacatcagcc tggacaccct ggtgcctgcc aagtttgagt tcattgtccg caggaaaggc
 721 ttccacaagg tcatggaggg catccacaag gccatcgagc tgggctacaa ccctgtgaag
 781 gtgaactgtg tggtgatgcg aggccttaac gaggatgaac tcctggactt tgcggccttg
 841 actgagggcc tccccctgga tgtgcgcttc atagagtata tgccctttga tggcaacaag
 901 tggaacttca agaagatggt cagctataag gagatgctag acactgtccg gcagcagtgg
 961 ccagagctgg agaaggtgcc agaggaggaa tccagcacag ccaaggcctt taaaatccct
1021 ggcttccaag gccagatcag cttcatcaca tccatgtctg agcatttctg tgggacctgc
1081 aaccgcctgc gaatcacagc tgatgggaac ctcaaggtct gcctctttgg aaactctgag
1141 gtatccctgc gggatcacct gcgagctggg gcctctgagc aggagctgct gagaatcatt
1201 ggggctgctg tgggcaggaa gaagcggcag catgcaggca tgttcagtat ttcccagatg
1261 aagaaccggc ccatgatcct catcaagtta tttttgatgt tccccaattc cccaccagcc
```

```
1321 aatccaagca ttttctcctg ggacccgctc catgttcagg gtctaagacc cagaatgagt
1381 ttctccagcc aggtggccac tttatggaaa ggatgcaggg tcccccagac ccctcctcta
1441 gcccagcagc ggctggggtc tggctccttt cagagacact acacttcccg tgcagactca
1501 gatgccaact caaagtgcct tagcccaggt tcctgggctt ctgctgcccc ctcaggaccc
1561 cagctaacct cagaacaact aactcatgtg gactcggaag gacgggcagc tatggtagat
1621 gtgggcagga agccagacac agagcgggtg gctgtggctt cagccgtggt cctcctggga
1681 ccggtagcct tcaagcttgt ccagcagaac cagctcaaga aaggagatgc cctagtggtg
1741 gcccagctgg ctggagtcca ggcagccaag gtgaccagcc agctgatccc tctgtgccac
1801 cacgtggccc tgagccacat ccaggtgcag ctggagctgg acagcacacg ccatgccgtg
1861 aagatccagg catcttgccg ggctcggggc cccaccgggg tggagatgga ggccctgacc
1921 tctgctgcag tggccgccct caccctgtat gacatgtgca aggctgtcag cagggacatc
1981 gtgttggagg agatcaagct cattagcaag actggtggtc agcgggggga cttccatcgg
2041 gcttagcacc tgcccttctc acccatggcc cacccaggcc tggagctggg atgcaatgta
2101 ggctgaggga aagacgtcag gttcctttaa tcacagtcac tgtttgttta ccttgagcag
2161 taaacccgaa gtcagcctgc tctactacta acaaacaggc ctgctgctag atgatctcta
2221 atgaccaatg gggcttcctt tctataggga ggataccagc aggcccttaa gccttccagg
2281 acactaaggt cgtgggagcg ggactgcaac aagcaatgcc agataactga gaaatcatgt
2341 tctttgtgga ctatttcaga caaccaggtt ccgacagtcc agcccagaac ttttccttct
2401 cattttgggt tttctcttct cctgctttcc tggggagaga ttaagcgctc attaagcaga
2461 ggagcccact ttgaggagag caaagcacaa gcttgcctga agaatggatc ccaacttctc
2521 cccggcagct ctgcctccct aagtctgtga agccgcagcc ctgccctgtc ctgtcctgtc
2581 ctgacttcat ctctccttct gcccaagtct gtgtcccatc agacttgcag cctttcagct
2641 taacagttgc ccggtcctgc tggccccttt tcctctggcc cccctcttct gaaacaggat
2701 gtgcacacat gggccatagc cctaaggact cctgccagac cacacagccc acacctggcc
2761 ctgttcacgg ctgttccacg cacccctctt tattctggag catatcaggg aaagaaaagt
2821 tgatgataga ttgccttcac cctcacagcg cacaaataaa gctacgatgc caactttgaa
2881 a
```

FIGURE 18G: Predicted amino acid sequence of human MOCS1, isoform 2 protein (SEQ ID NO:22)

```
  1 mwkswklrtd vrvregaggs pcassqpgsr gpcflpglss qevsrrrqfl rehaapfsaf
 61 ltdsfgrqhs ylrisltekc nlrcqycmpe egvpltpkan lltteeiltl arlfvkegid
121 kirltggepl irpdvvdiva qlqrleglrt igvttnginl arllpqlqka glsainisld
181 tlvpakfefi vrrkgfhkvm egihkaielg ynpvkvncvv mrglnedell dfaalteglp
241 ldvrfieymp fdgnkwnfkk mvsykemldt vrqqwpelek vpeeesstak afkipgfqgq
301 isfitsmseh fcgtcnrlri tadgnlkvcl fgnsevslrd hlragaseqe llriigaavg
361 rkkrqhagmf sisqmknrpm iliklflmfp nsppanpsif swdplhvqgl rprmsfssqv
421 atlwkgcrvp qtpplaqqrl gsgsfqrhyt sradsdansk clspgswasa apsgpqltse
481 qlthvdsegr aamvdvgrkp dtervavasa vvllgpvafk lvqqnqlkkg dalvvaqlag
541 vqaakvtsql iplchhvals hiqvqlelds trhavkiqas crargptgve mealtsaava
601 altlydmcka vsrdivleei klisktgggr gdfhra
```

FIGURE 18H: Predicted nucleotide sequence of human MOCS1, isoform 3 protein (SEQ ID NO:23)

```
   1 gccagaaatc ttcccagtag agatcaccat ccgcccccga cccccaagaa tacttaaggg
  61 gtgggtcctt cccatcaagc tgatttctca acgagaggga caatcccagc ttccccaaca
 121 ttgcagagcc caaacatgtg gaagagttgg aagctccgca cagatgtcag agtaagggag
 181 ggggcaggcg gttctccttg tgcctcttcc cagcccggta gcaggggccc atgcttcctc
 241 cctggtctgt cctcgcagga ggtgtccagg cggaggcagt tcctgcggga gcatgcggcc
 301 cccttctccg ccttcctcac agacagcttc ggccggcagc acagctacct gcggatctcc
 361 ctcacagaga agtgcaacct cagatgtcag tactgcatgc ccgaggaggg ggtcccgctg
 421 acccccaaag ccaacctgct gaccacagag gagatcctga ccctcgcccg gctctttgtg
 481 aaggaaggca tcgacaagat ccggctcaca ggtggagagc cgcttatccg gccggacgtg
```

```
 541 gtggacattg tggcccagct ccagcggctg gaagggctga gaaccatagg tgttaccacc
 601 aatggcatca acctggcccg gctactgccc cagcttcaga aggctggtct cagtgccatc
 661 aacatcagcc tggacaccct ggtgcctgcc aagtttgagt tcattgtccg caggaaaggc
 721 ttccacaagg tcatggaggg catccacaag gccatcgagc tgggctacaa ccctgtgaag
 781 gtgaactgtg tggtgatgcg aggccttaac gaggatgaac tcctggactt tgcggccttg
 841 actgagggcc tccccctgga tgtgcgcttc atagagtata tgccctttga tggcaacaag
 901 tggaacttca agaagatggt cagctataag gagatgctag acactgtccg gcagcagtgg
 961 ccagagctgg agaaggtgcc agaggaggaa tccagcacag ccaaggcctt taaaatccct
1021 ggcttccaag gccagatcag cttcatcaca tccatgtctg agcatttctg tgggacctgc
1081 aaccgcctgc gaatcacagc tgatgggaac ctcaaggtct gcctctttgg aaactctgag
1141 gtatccctgc gggatcacct gcgagctggg gcctctgagc aggagctgct gagaatcatt
1201 ggggctgctg tgggcaggaa gaagcggcag catgcaaagt tatttttgat gttccccaat
1261 tccccaccag ccaatccaag cattttctcc tgggacccgc tccatgttca gggtctaaga
1321 cccagaatga gtttctccag ccaggtggcc actttatgga aaggatgcag ggtcccccag
1381 acccctcctc tagcccagca gcggctgggg tctggctcct ttcagagaca ctacacttcc
1441 cgtgcagact cagatgccaa ctcaaagtgc cttagcccag gttcctgggc ttctgctgcc
1501 ccctcaggac cccagctaac ctcagaacaa ctaactcatg tggactcgga aggacgggca
1561 gctatggtag atgtgggcag gaagccagac acagagcggg tggctgtggc ttcagccgtg
1621 gtcctcctgg gaccggtagc cttcaagctt gtccagcaga accagctcaa gaaaggagat
1681 gccctagtgg tggcccagct ggctggagtc caggcagcca aggtgaccag ccagctgatc
1741 cctctgtgcc accacgtggc cctgagccac atccaggtgc agctggagct ggacagcaca
1801 cgccatgccg tgaagatcca ggcatcttgc cgggctcggg gccccaccgg ggtggagatg
1861 gaggccctga cctctgctgc agtggccgcc ctcaccctgt atgacatgtg caaggctgtc
1921 agcagggaca tcgtgttgga ggagatcaag ctcattagca agactggtgg tcagcggggg
1981 gacttccatc gggcttagca cctgcccttc tcacccatgg cccacccagg cctggagctg
2041 ggatgcaatg taggctgagg gaaagacgtc aggttccttt aatcacagtc actgtttgtt
2101 taccttgagc agtaaacccg aagtcagcct gctctactac taacaaacag gcctgctgct
2161 agatgatctc taatgaccaa tggggcttcc tttctatagg gaggatacca gcaggccctt
2221 aagccttcca ggacactaag gtcgtgggag cgggactgca acaagcaatg ccagataact
2281 gagaaatcat gttctttgtg gactatttca gacaaccagg ttccgacagt ccagcccaga
2341 acttttcctt ctcattttgg gttttctctt ctcctgcttt cctggggaga gattaagcgc
2401 tcattaagca gaggagccca ctttgaggag agcaaagcac aagcttgcct gaagaatgga
2461 tcccaacttc tccccggcag ctctgcctcc ctaagtctgt gaagccgcag ccctgccctg
2521 tcctgtcctg tcctgacttc atctctcctt ctgcccaagt ctgtgtccca tcagacttgc
2581 agcctttcag cttaacagtt gcccggtcct gctggcccct tttcctctgg ccccctctt
2641 ctgaaacagg atgtgcacac atgggccata gccctaagga ctcctgccag accacacagc
2701 ccacacctgg ccctgttcac ggctgttcca cgcacccctc tttattctgg agcatatcag
2761 ggaaagaaaa gttgatgata gattgccttc accctcacag cgcacaaata aagctacgat
2821 gccaactttg aaa
```

FIGURE 18I: Predicted amino acid sequence of human MOCS1, isoform 3 protein (SEQ ID NO:24)

```
1 mwkswklrtd vrvregaggs pcassqpgsr gpcflpglss qevsrrrqfl rehaapfsaf
 61 ltdsfgrqhs ylrisltekc nlrcqycmpe egvpltpkan lltteeiltl arlfvkegid
121 kirltggepl irpdvvdiva qlqrleglrt igvttnginl arllpqlqka glsainisld
181 tlvpakfefi vrrkgfhkvm egihkaielg ynpvkvncvv mrglnedell dfaalteglp
241 ldvrfieymp fdgnkwnfkk mvsykemldt vrqqwpelek vpeeesstak afkipgfqgq
301 isfitsmseh fcgtcnrlri tadgnlkvcl fgnsevslrd hlragaseqe llriigaavg
361 rkkrqhaklf lmfpnsppan psifswdplh vqglrprmsf ssqvatlwkg crvpqtppla
421 qqrlgsgsfq rhytsradsd anskclspgs wasaapsgpq ltseqlthvd segraamvdv
481 grkpdterva vasavvllgp vafklvqqnq lkkgdalvva qlagvqaakv tsqliplchh
541 valshiqvql eldstrhavk iqascrargp tgvemealts aavaaltlyd mckavsrdiv
601 leeikliskt ggqrgdfhra
```

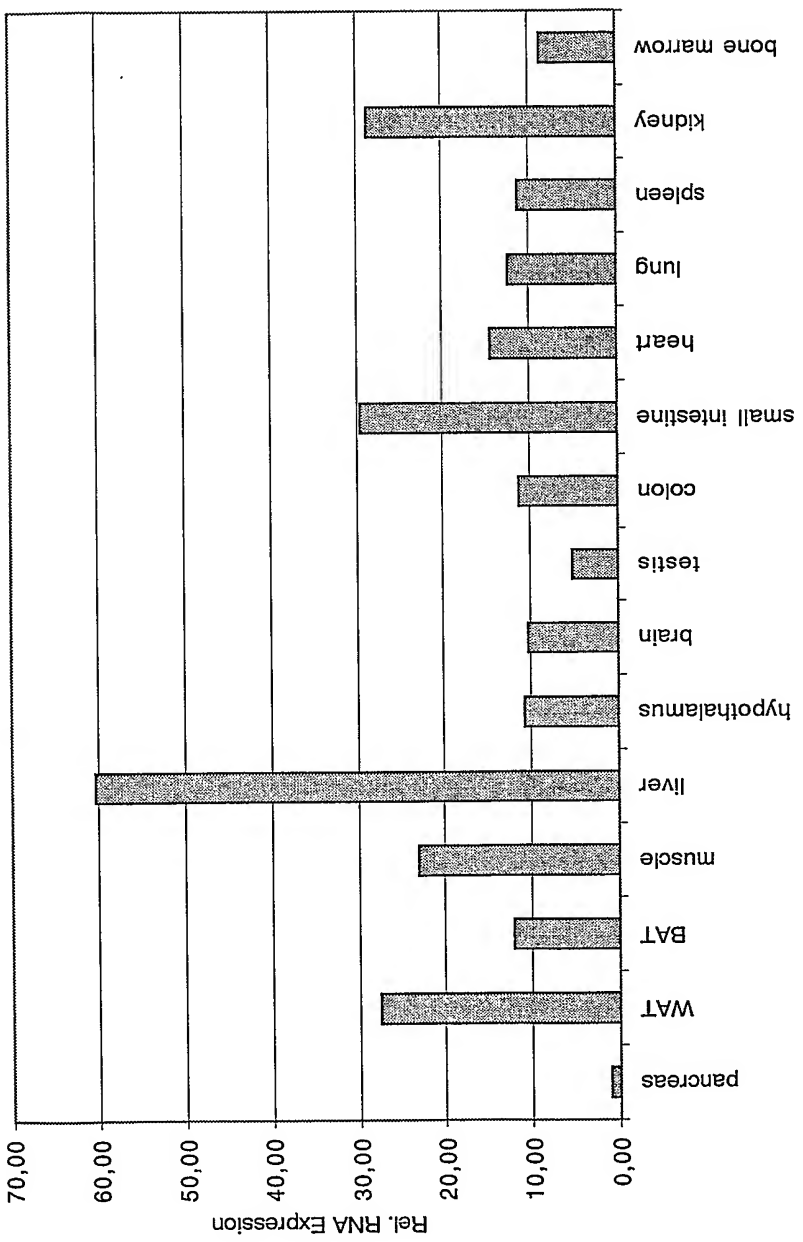
FIGURE 19. CLUSTAL W (1.82) Protein Sequence Alignment Analysis

```
Mocs1-2 Hs   MWKSWKLRTDVRVREGAGGSPCASSQPGSR-----GPCFLPGLSSQEVSRRRQFLREHAA
Mocs1-3 Hs   MWKSWKLRTDVRVREGAGGSPCASSQPGSR-----GPCFLPGLSSQEVSRRRQFLREHAA
Mocs1-1 Hs   MWKSWKLRTDVRVREGAGGSPCASSQPGSR-----GPCFLPGLSSQEVSRRRQFLREHAA
Mocs1 Hs     -MAARPLSRMLRRLLRSSARSCSSGAPVTQP----CPGESARAASEEVSRRRQFLREHAA
Mocs1-PA Dm MRLLARHAIRLLGQENSAGEVASLSRGAIRLKATTGYLNLATASVQPLEPEKQVLRKNSP
Mocs1 PC Dm MRLLARHAIRLLGQENSAGEVASLSRGAIRLKATTGYLNLATASVQPLEPEKQVLRKNSP

Mocs1-2 Hs   PFSAFLTDSFGRQHSYLRISLTEKCNLRCQYCMPEEGVPLTPKANLLTTEEILTLARLFV
Mocs1-3 Hs   PFSAFLTDSFGRQHSYLRISLTEKCNLRCQYCMPEEGVPLTPKANLLTTEEILTLARLFV
Mocs1-1 Hs   PFSAFLTDSFGRQHSYLRISLTEKCNLRCQYCMPEEGVPLTPKANLLTTEEILTLARLFV
Mocs1 Hs     PFSAFLTDSFGRQHSYLRISLTEKCNLRCQYCMPEEGVPLTPKANLLTTEEILTLARLFV
Mocs1-PA Dm -----LTDSFGRHHTYLRISLTERCNLRCDYCMPAEGVPLQPKNKLLTTEEILRLARIFV
Mocs1 PC Dm -----LTDSFGRHHTYLRISLTERCNLRCDYCMPAEGVPLQPKNKLLTTEEILRLARIFV

Mocs1-2 Hs   KEGIDKIRLTGGEPLIRPDVVDIVAQLQRLEGLRTIGVTTNGINLARLLPQLQKAGLSAI
Mocs1-3 Hs   KEGIDKIRLTGGEPLIRPDVVDIVAQLQRLEGLRTIGVTTNGINLARLLPQLQKAGLSAI
Mocs1-1 Hs   KEGIDKIRLTGGEPLIRPDVVDIVAQLQRLEGLRTIGVTTNGINLARLLPQLQKAGLSAI
Mocs1 Hs     KEGIDKIRLTGGEPLIRPDVVDIVAQLQRLEGLRTIGVTTNGINLARLLPQLQKAGLSAI
Mocs1-PA Dm EQGVRKIRLTGGEPTVRRDIVEIVAQMKALPELEQIGITTNGLVLTRLLLPLQRAGLDNL
Mocs1 PC Dm EQGVRKIRLTGGEPTVRRDIVEIVAQMKALPELEQIGITTNGLVLTRLLLPLQRAGLDNL

Mocs1-2 Hs   NISLDTLVPAKFEFIVRRKGFHKVMEGIHKAIELGYNPVKVNCVVMRGLNEDELLDFAAL
Mocs1-3 Hs   NISLDTLVPAKFEFIVRRKGFHKVMEGIHKAIELGYNPVKVNCVVMRGLNEDELLDFAAL
Mocs1-1 Hs   NISLDTLVPAKFEFIVRRKGFHKVMEGIHKAIELGYNPVKVNCVVMRGLNEDELLDFAAL
Mocs1 Hs     NISLDTLVPAKFEFIVRRKGFHKVMEGIHKAIELGYNPVKVNCVVMRGLNEDELLDFAAL
Mocs1-PA Dm NISLDTLKRDRFEKITRRKGWERVIAGIDLAVQLGYRP-KVNCVLMRDFNEDEICDFVEF
Mocs1 PC Dm NISLDTLKRDRFEKITRRKGWERVIAGIDLAVQLGYRP-KVNCVLMRDFNEDEICDFVEF

Mocs1-2 Hs   TEGLPLDVRFIEYMPFDGNKWNFKKMVSYKEMLDTVRQQWPELEKVPEEESSTAKAFKIP
Mocs1-3 Hs   TEGLPLDVRFIEYMPFDGNKWNFKKMVSYKEMLDTVRQQWPELEKVPEEESSTAKAFKIP
Mocs1-1 Hs   TEGLPLDVRFIEYMPFDGNKWNFKKMVSYKEMLDTVRQQWPELEKVPEEESSTAKAFKIP
Mocs1 Hs     TEGHPLDVRFIEYMPFDGNKWNFKKMVSYKEMLDTVRQQWPELEKVPEEESSTAKAFKIP
Mocs1-PA Dm TRNRPVDVRFIEYMPFSGNKWHTERLISYKDTLQIIRQRWPDFKALPNGPNDTSKAYAVP
Mocs1 PC Dm TRNRPVDVRFIEYMPFSGNKWHTERLISYKDTLQIIRQRWPDFKALPNGPNDTSKAYAVP

Mocs1-2 Hs   GFQGQISFITSMSEHFCGTCNRLRITADGNLKVCLFGNSEVSLRDHLR-AGASEQELLRI
Mocs1-3 Hs   GFQGQISFITSMSEHFCGTCNRLRITADGNLKVCLFGNSEVSLRDHLR-AGASEQELLRI
Mocs1-1 Hs   GFQGQISFITSMSEHFCGTCNRLRITADGNLKVCLFGNSEVSLRDHLR-AGASEQELLRI
Mocs1 Hs     GFQGQISFITSMSEHFCGTCNRLRITADGNLKVCLFGNSEVSLRDHLR-AGASEQELLRI
Mocs1-PA Dm GFKGQVGFITSMTEHFCGTCNRLRLTADGNIKVCLFGNKEFSLRDAMRDESVSEEQLVDL
Mocs1 PC Dm GFKGQVGFITSMTEHFCGTCNRLRLTADGNIKVCLFGNKEFSLRDAMRDESVSEEQLVDL

Mocs1-2 Hs   IGAAVGRKKRQHAGMFSISQMKNRPMILIKLFLMFPNSPPANPSIFSWDPLHVQGLRPRM
Mocs1-3 Hs   IGAAVGRKKRQHA----------------KLFLMFPNSPPANPSIFSWDPLHVQGLRPRM
Mocs1-1 Hs   IGAAVGRKKRQHAG-------------------MFSISQMKN--------------RPMI
Mocs1 Hs     IGAAVGRKKRQHAG-------------------MFSISQMKN--------------RPMI
Mocs1-PA Dm IGAAVQRKKKQHAG-----------------------------------------------ML
Mocs1 PC Dm IGAAVQRKKKQHADA-----------------APRLHHHLHPYSYHHAYHTSRLQLQAR

Mocs1-2 Hs   SFSSQVATLWKGCRVPQTPPLAQQRLGSGSFQRHYTSRADSDANSKCLSPGSWASAAPSG
Mocs1-3 Hs   SFSSQVATLWKGCRVPQTPPLAQQRLGSGSFQRHYTSRADSDANSKCLSPGSWASAAPSG
Mocs1-1 Hs   LIGG------------------------------------------------------------
Mocs1 Hs     LIGG------------------------------------------------------------
Mocs1-PA Dm NLS-------------------------------------------------------------
```

```
Mocs1 PC Dm NYS-----------------------------------------------------

Mocs1-2 Hs  PQLTSEQLTHVDSEGRAAMVDVGRKPDTERVAVASAVVLLGPVAFKLVQQNQLKKGDALV
Mocs1-3 Hs  PQLTSEQLTHVDSEGRAAMVDVGRKPDTERVAVASAVVLLGPVAFKLVQQNQLKKGDALV
Mocs1-1 Hs  ------------------------------------------------------------
Mocs1 Hs    ------------------------------------------------------------
Mocs1-PA Dm ------QMENR------PMILIGG-------------------------------------
Mocs1 PC Dm ------QLTHVDGQGKAQMVDVGAKPSTTRLARAEATVQVGEKLTQLIADNQVAKGDVLT

Mocs1-2 Hs  VAQLAGVQAAKVTSQLIPLCHHVALSHIQVQLELDSTRHAVKIQASCRARGPTGVEMEAL
Mocs1-3 Hs  VAQLAGVQAAKVTSQLIPLCHHVALSHIQVQLELDSTRHAVKIQASCRARGPTGVEMEAL
Mocs1-1 Hs  ------------------------------------------------------------
Mocs1 Hs    ------------------------------------------------------------
Mocs1-PA Dm ------------------------------------------------------------
Mocs1 PC Dm VAQIAGIMGAKRTAELIPLCHNISLSSVKVQATLLKTEQSVRLEATVRCSGQTGVEMEAL

Mocs1-2 Hs  TSAAVAALTLYDMCKAVSRDIVLEEIKLISKTGGQRGDFHRA----------
Mocs1-3 Hs  TSAAVAALTLYDMCKAVSRDIVLEEIKLISKTGGQRGDFHRA----------
Mocs1-1 Hs  ------------------------------------------------
Mocs1 Hs    ------------------------------------------------
Mocs1-PA Dm ------------------------------------------------
Mocs1 PC Dm TAVSVAALTVYDMCKAVSHDICITNVRLLSKSGGKRDFQREEPQNGIVTEVE
```

FIGURE 20. Expression of Mocs in mammalian tissues

FIGURE 20A. Real-time PCR analysis of Mocs expression in wildtype mouse tissues



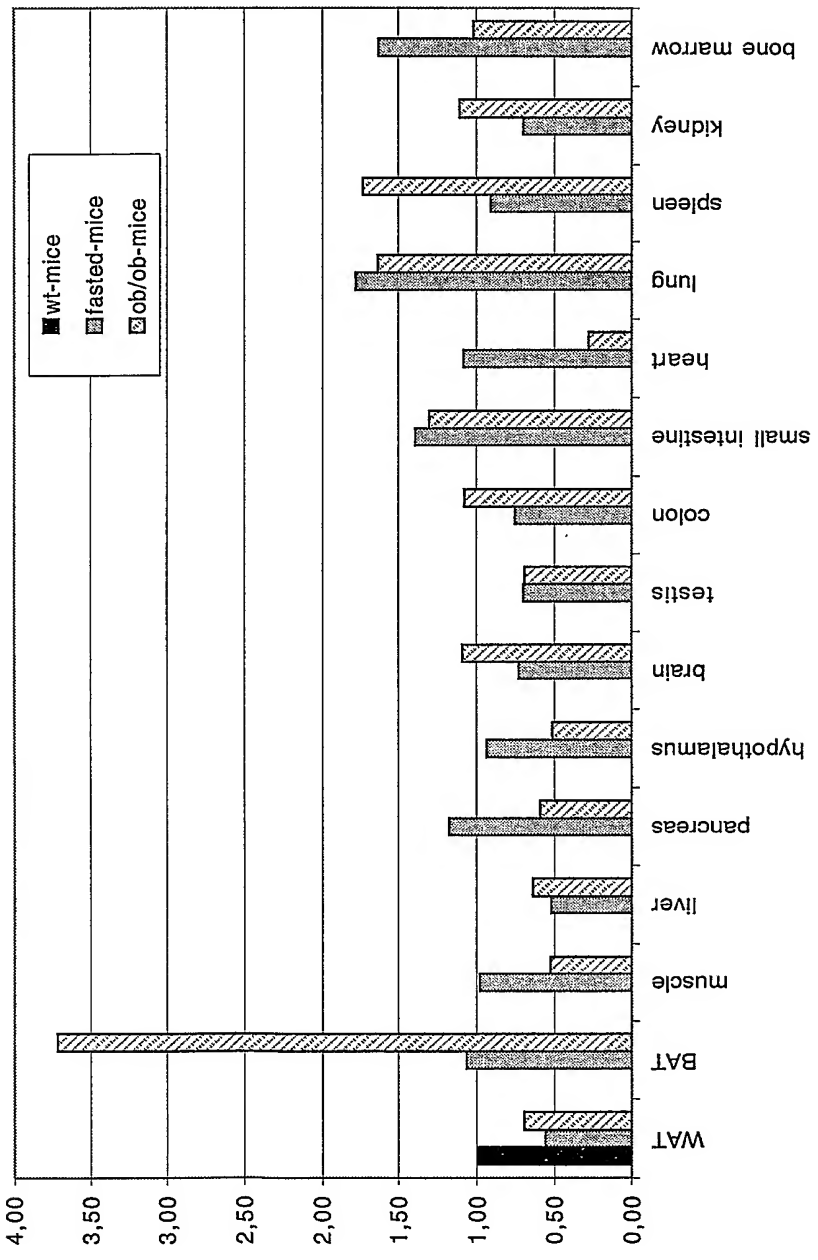


FIGURE 20B. Real-time PCR analysis of Mocs expression in different mouse models

FIGURE 21. Triglyceride content of a *peanut* (pnut; Gadfly Accession Number CG8705) mutant

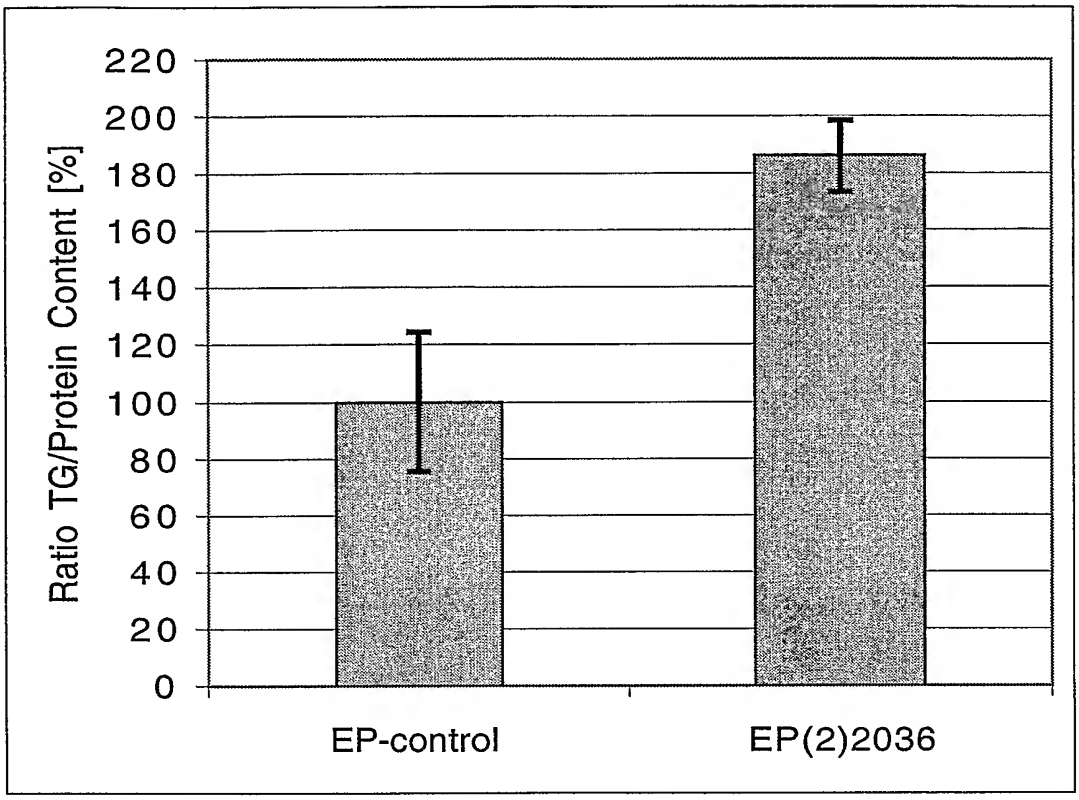


FIGURE 22. Molecular organisation of the *peanut* gene (GadFly Accession Number CG8705)

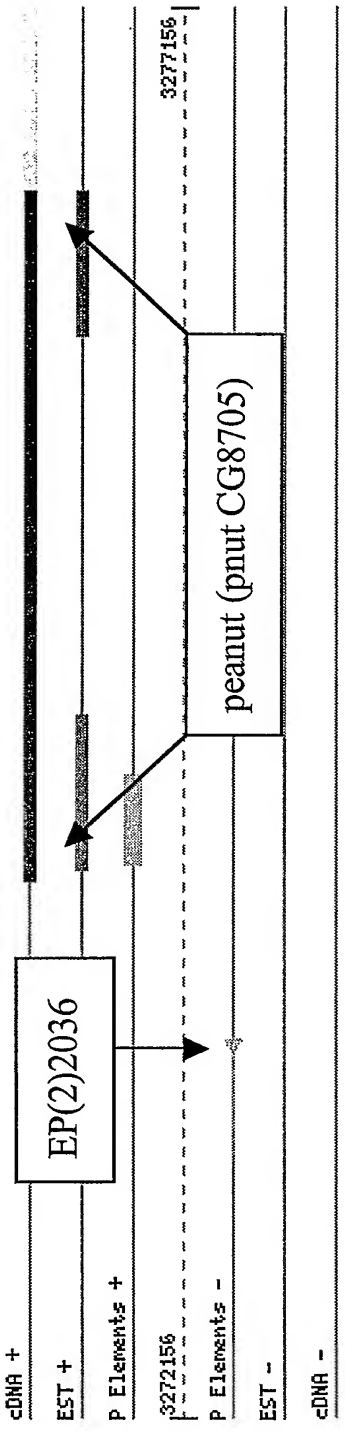


FIGURE 23: HUMAN HOMOLOG OF CG8705 (peanut)

FIGURE 23A. BLASTP search results for peanut (Gadfly Accession Number CG8705)

Homology to human CDC10 protein (gene ref NM_001788; protein ref NP_001779.1)

```
>ref|NP_001779.1| (NM_001788) cell division cycle 10; cell division cycle
10 (homolog to CDC10 of S.cerevisiae); cell division cycle 10 (homologous
to CDC10 of S. cerevisiae); CDC10 (cell division cycle 10, S. cerevisiae,
homolog); CDC10 protein homolog [Homo sapiens]
sp|Q16181|SEP7_HUMAN SEPTIN 7 (CDC10 PROTEIN HOMOLOG)
pir||JC2352 hCDC10 protein - human
gb|AAB31337.1| (S72008) CDC10 homolog [Homo sapiens]
Length = 418

Score =  548 bits (1398), Expect = e-155
Identities = 273/419 (65%), Positives = 331/419 (78%), Gaps = 9/419 (2%)

Query: 113 RQKPMEIAGYVGFANLPNQVYRKAVKRGFEFTLMVVGASGLGKSTLINSMFLSDIYNAEQ 172
           +QK +E  GYVGFANLPNQVYRK+VKRGFEFTLMVVG SGLGKSTLINS+FL+D+Y+ E
Sbjct: 4   QQKNLE--GYVGFANLPNQVYRKSVKRGFEFTLMVVGESGLGKSTLINSLFLTDLYSPE- 60

Query: 173 YPGPSLRKKKTVAVEATKVMLKENGVNLTLTVVDTPGFGDAVDNSNCWVPILEYVDSKYE 232
           YPGPS R KKTV VE +KV++KE GV L LT+VDTPGFGDAVDNSNCW P+++Y+DSK+E
Sbjct: 61  YPGPSHRIKKTVQVEQSKVLIKEGGVQLLLTIVDTPGFGDAVDNSNCWQPVIDYIDSKFE 120

Query: 233 EYLTAESRVVYRKTISDSRVHCCLYFIAPSGHGLLPLDIACMQSLSDKVNLVPVIAKADTM 292
           +YL AESRV R+ + D+RV CCLYFIAPSGHGL PLDI  M+ L +KVN++P+IAKADT+
Sbjct: 121 DYLNAESRVNRRQMPDNRVQCCLYFIAPSGHGLKPLDIEFMKRLHEKVNIIPLIAKADTL 180

Query: 293 TPDEVHLFKKQILNEIAQHKIKIYDFPATLEDAAEEAKTTQNLRSRVPFAVVGANTIIEQ 352
           TP+E   FKKQI+ EI +HKIKIY+FP T  D  EE K  + ++ R+P AVVG+NTIIE
Sbjct: 181 TPEECQQFKKQIMKEIQEHKIKIYEFPET--DDEEENKLVKKIKDRLPLAVVGSNTIIEV 238

Query: 353 DGKKVRGRRYPWGLVEVENLTHCDFIALRNMVIRTHLQDLKDVTNNVHYENYRCRKLSEL 412
           +GK+VRGR+YPWG+ EVEN  HCDF  LRNM IRTH+QDLKDVTNNVHYENYR RKL+ +
Sbjct: 239 NGKRVRGRQYPWGIAEVENGEHCDFTILRNMKIRTHMQDLKDVTNNVHYENYRSRKLAAV 298

Query: 413 ---GLVDGKAR-LSNKNPLTQMEEEKREHEQKMKKMEAEMEQVFDMKVKEKMQKLRDSEL 468
              G+ + K +    K+PL QMEEE+REH  KMKKME EMEQVF+MKVKEK+QKL+DSE
Sbjct: 299 TYNGVDNNKNKGQLTKSPLAQMEEERREHVAKMKKMEMEMEQVFEMKVKEKVQKLKDSEA 358

Query: 469 ELARRHEERKKALELQIRELEEKRREFEREKKEWEDVNHVTLEELKRRSLGANSSTDNV 527
           EL RRHE+ KK LE Q +ELEEKRR+FE EK  WE    +  ++   R+L  N     +
Sbjct: 359 ELQRRHEQMKKNLEAQHKELEEKRRQFEDEKANWEAQQRILEQQNSSRTLEKNKKKGKI 417
```

Homology to human Septin 7 (gene ref XM_011595; protein ref XP_011595.4)

```
>ref| | (XM_011595) similar to SEPTIN 7 (CDC10 PROTEIN HOMOLOG) (H.
sapiens) [Homo sapiens], Length = 384

Score =  498 bits (1268), Expect = e-139
Identities = 246/386 (63%), Positives = 302/386 (77%), Gaps = 7/386 (1%)
```

```
Query: 146 MVVGASGLGKSTLINSMFLSDIYNAEQYPGPSLRKKKTVAVEATKVMLKENGVNLTLTVV 205
           MVVG SGLGKSTLINS+FL+D+Y+ E YPGPS R KKTV VE +KV++KE GV L LT+V
Sbjct: 1   MVVGESGLGKSTLINSLFLTDLYSPE-YPGPSHRIKKTVQVEQSKVLIKEGGVQLLLTIV 59

Query: 206 DTPGFGDAVDNSNCWVPILEYVDSKYEEYLTAESRVYRKTISDSRVHCCLYFIAPSGHGL 265
           DTPGFGDAVDNSNCW P+++Y+DSK+E+YL AESRV R+ + D+RV CCLYFIAPSGHGL
Sbjct: 60  DTPGFGDAVDNSNCWQPVIDYIDSKFEDYLNAESRVNRRQMPDNRVQCCLYFIAPSGHGL 119

Query: 266 LPLDIACMQSLSDKVNLVPVIAKADTMTPDEVHLFKKQILNEIAQHKIKIYDFPATLEDA 325
            PLDI  M+ L +KVN++P+IAKADT+TP+E   FKKQI+ EI +HKIKIY+FP T  D
Sbjct: 120 KPLDIEFMKRLHEKVNIIPLIAKADTLTPEECQQFKKQIMKEIQEHKIKIYEFPET--DD 177

Query: 326 AEEAKTTQNLRSRVPFAVVGANTIIEQDGKKVRGRRYPWGLVEVENLTHCDFIALRNMVI 385
            EE K  + ++ R+P AVVG+NTIIE +GK+VRGR+YPWG+ EVEN  HCDF  LRNM+I
Sbjct: 178 EEENKLVKKIKDRLPLAVVGSNTIIEVNGKRVRGRQYPWGVAEVENGEHCDFTILRNMLI 237

Query: 386 RTHLQDLKDVTNNVHYENYRCRKLSEL---GLVDGKAR-LSNKNPLTQMEEEKREHEQKM 441
           RTH+QDLKDVTNNVHYENYR RKL+ +   G+ + K +     K+PL QMEEE+REH  KM
Sbjct: 238 RTHMQDLKDVTNNVHYENYRSRKLAAVTYNGVDNNKNKGQLTKSPLAQMEEERREHVAKM 297

Query: 442 KKMEAEMEQVFDMKVKEKMQKLRDSELELARRHEERKKALELQIRELEEKRREFEREKKE 501
           KKME EMEQVF+MKVKEK+QKL+DSE EL RRHE+ KK LE Q +ELEEKRR+FE EK
Sbjct: 298 KKMEMEMEQVFEMKVKEKVQKLKDSEAELQRRHEQMKKNLEAQHKELEEKRRQFEDEKAN 357

Query: 502 WEDVNHVTLEELKRRSLGANSSTDNV 527
           WE    +  ++   R+L  N     +
Sbjct: 358 WEAQQRILEQQNSSRTLEKNKKKGKI 383
```

FIGURE 23B: Predicted nucleotide sequence encoding human CDC10 cell division cycle 10 homolog (SEQ ID NO:25)

```
   1 agtgcgagat ccgctgctgc tgaggagagg agcgtcaaca gcagcaccat ggtagctcaa
  61 cagaagaacc ttgaaggcta tgtgggattt gccaatctcc caaatcaagt atacagaaaa
 121 tcggtgaaga gaggttttga attcacgctt atggtagtgg gtgaatctgg attgggaaag
 181 tcgacattaa tcaactcatt attcctcaca gatttgtatt ctccagagta tccaggtcct
 241 tctcatagaa ttaaaaagac tgtacaggtg gaacaatcca aagttttaat caaagaaggt
 301 ggtgttcagt tgctgctcac aatagttgat accccaggat ttggagatgc agtggataat
 361 agtaattgct ggcagcctgt tatcgactac attgatagta aatttgagga ctacctaaat
 421 gcagaatcac gagtgaacag acgtcagatg cctgataaca gggtgcagtg ttgtttatac
 481 ttcattgctc cttcaggaca tggacttaaa ccattggata ttgagtttat gaagcgtttg
 541 catgaaaaag tgaatatcat cccacttatt gccaaagcag acacactcac accagaggaa
 601 tgccaacagt ttaaaaaaca gataatgaaa gaaatccaag aacataaaat taaaatatac
 661 gaatttccag aaacagatga tgaagaagaa aataaacttg ttaaaaagat aaaggaccgt
 721 ttacctcttg ctgtggtagg tagtaatact atcattgaag ttaatggcaa aagggtcaga
 781 ggaaggcagt atccttgggg tattgctgaa gttgaaaatg gtgaacattg tgattttaca
 841 atcctaagaa atatgaagat aagaacacac atgcaggact tgaaagatgt tactaataat
 901 gtccactatg agaactacag aagcagaaaa cttgcagctg tgacttataa tggagttgat
 961 aacaacaaga ataaagggca gctgactaag agccctctgg cacaaatgga agaagaaaga
1021 agggagcatg tagctaaaat gaagaagatg gagatggaga tggagcaggt gtttgagatg
1081 aaggtcaaag aaaaagttca aaaactgaag gactctgaag ctgagctcca gcggcgccat
1141 gagcaaatga aaaagaattt ggaagcacag cacaaagaat tggaggaaaa acgtcgtcag
1201 ttcgaggatg agaaagcaaa ctgggaagct caacaacgta ttttagaaca acagaactct
1261 tcaagaacct tggaaaagaa caagaagaaa gggaagatct tttaaactct ctattgacca
```

```
1321 ccagttaacg tattagttgc caatatgcca gcttggacat cagtgtttgt tggatccgtt
1381 tgaccaattt gcaccagttt tatccataat gatggattta acagcatgac aaaaattatt
1441 ttttttttg ttcttgatgg agattaagat gccttgaatt gtctagggtg ttctgtactt
1501 agaaagtaag agctctaagt acctttccta cattttcttt ttttattaaa cagatatctt
1561 cagtttaatg caagagaaca ttttactgtt gtacaatcat gttctggtgg tttgattgtt
1621 tacaggatat tccaaaataa aaggactctg gaagattttc attgaggata aattgccata
1681 atatgatgca aactgtgctt ctctatgata attacaatac aaaggttcca ttcagtgcag
1741 catatacaat aatgtaattt agtctaacac agttgaccct atttttgac acttccattg
1801 tttaaaaata cacatggaaa aaaaaaaacc ctatatgctt actgtgcacc tagagctttt
1861 ttataacaac gtctttttgt ttgtttgttt tggattcttt aaatatatat tattctcatt
1921 tagtgccctc tttagccaga atctcattac tgcttcattt ttgtaataac atttaattta
1981 gatattttcc atcatattgg cactgctaaa atagaatata gcatctttca tatggtagga
2041 accaacaagg aaactttcct ttaactccct ttttacactt tatggtaagt agcaggggg
2101 gaaatgcatt tatagatcat ttctaggcaa aattgtgaag ctaatgacca acctgtttct
2161 acctatatgc agtctcttta ttttactaga aatgggaatc atggcctctt gaagagaaaa
2221 aagtcaccat tctgcattta gctgtattca tatattgcta tttctgtatt ttttgtttgt
2281 attgtaaaaa attcacataa taaacgatgg ttgtgatgt
```

FIGURE 23C: Predicted amino acid sequence of human CDC10 cell division cycle 10 homolog (SEQ ID NO:26)

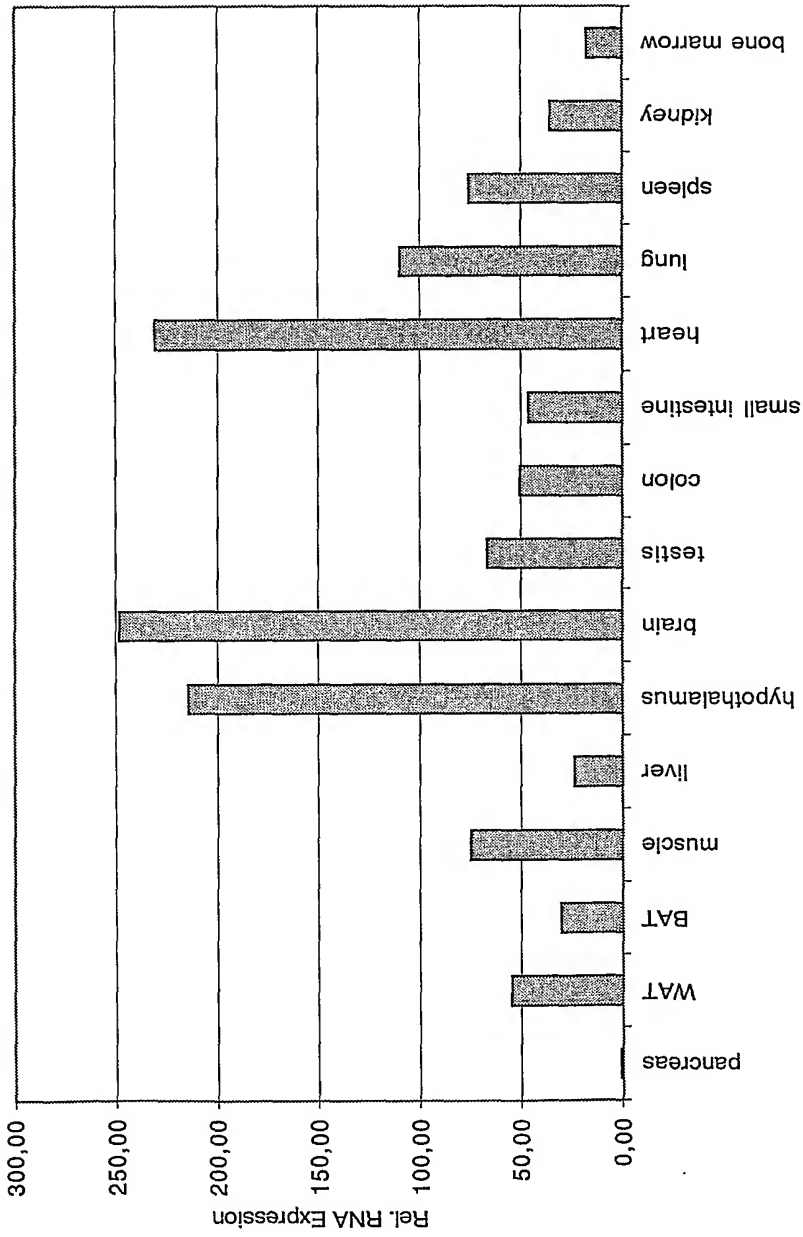
```
  1 mvaqqknleg yvgfanlpnq vyrksvkrgf eftlmvvges glgkstlins lfltdlyspe
 61 ypgpshrikk tvqveqskvl ikeggvqlll tivdtpgfgd avdnsncwqp vidyidskfe
121 dylnaesrvn rrqmpdnrvq cclyfiapsg hglkpldief mkrlhekvni ipliakadtl
181 tpeecqqfkk qimkeiqehk ikiyefpetd deeenklvkk ikdrlplavv gsntiievng
241 krvrgrqypw giaevengeh cdftilrnmk irthmqdlkd vtnnvhyeny rsrklaavty
301 ngvdnnknkg qltksplaqm eeerrehvak mkkmememeq vfemkvkekv qklkdseael
361 qrrheqmkkn leaqhkelee krrqfedeka nweaqqrile qqnssrtlek nkkkgkif
```

FIGURE 24. CLUSTAL W (1.7) Protein Sequence Alignment Analysis

```
XM_011595       ------------------------------------------------------------
NM_001788       ------------------------------------------------------------
pnut            MNSPRSNAVNGGSGGAISALPSTLAQLALRDKQQAASASASSATNGSSGSESLVGVGGRP


XM_011595       ------------------------------------------------------------
NM_001788       -----------------------------------------------MVAQQK--NLE
pnut            PNQPPSVPVAASGKLDTSSGGASNGDSNKLTHDLQEKEHQQAQKPQKPPLPVRQKPMEIA


XM_011595       -------------------------MVVGESGLGKSTLINSLFLTDLY-SPEYPGPSHRI
NM_001788       GYVGFANLPNQVYRKSVKRGFEFTLMVVGESGLGKSTLINSLFLTDLY-SPEYPGPSHRI
pnut            GYVGFANLPNQVYRKAVKRGFEFTLMVVGASGLGKSTLINSMFLSDIYNAEQYPGPSLRK
                                        **** ***********:**:*:* : :***** *

XM_011595       KKTVQVEQSKVLIKEGGVQLLLTIVDTPGFGDAVDNSNCWQPVIDYIDSKFEDYLNAESR
NM_001788       KKTVQVEQSKVLIKEGGVQLLLTIVDTPGFGDAVDNSNCWQPVIDYIDSKFEDYLNAESR
pnut            KKTVAVEATKVMLKENGVNLTLTVVDTPGFGDAVDNSNCWVPILEYVDSKYEEYLTAESR
                **** ** :**::**.**:* **:**************** *:::*:***:*:**.****

XM_011595       VNRRQMPDNRVQCCLYFIAPSGHGLKPLDIEFMKRLHEKVNIIPLIAKADTLTPEECQQF
NM_001788       VNRRQMPDNRVQCCLYFIAPSGHGLKPLDIEFMKRLHEKVNIIPLIAKADTLTPEECQQF
pnut            VYRKTISDNRVHCCLYFIAPSGHGLLPLDIACMQSLSDKVNLVPVIAKADTMTPDEVHLF
                * *: :.****:************** ****  *: * :***::*:******:**:* : *

XM_011595       KKQIMKEIQEHKIKIYEFPETDDE--EENKLVKKIKDRLPLAVVGSNTIIEVNGKRVRGR
NM_001788       KKQIMKEIQEHKIKIYEFPETDDE--EENKLVKKIKDRLPLAVVGSNTIIEVNGKRVRGR
pnut            KKQILNEIAQHKIKIYDFPATLEDAAEEAKTTQNLRSRVPFAVVGANTIIEQDGKKVRGR
                ****::** :******:** * ::  ** * .:::::.*:*:****:***** :**:****

XM_011595       QYPWGVAEVENGEHCDFTILRNMLIRTHMQDLKDVTNNVHYENYRSRKLAAVTYNGVDNN
NM_001788       QYPWGIAEVENGEHCDFTILRNMKIRTHMQDLKDVTNNVHYENYRSRKLAAVTYNGVDNN
pnut            RYPWGLVEVENLTHCDFIALRNMVIRTHLQDLKDVTNNVHYENYRCRKLSELGL--VDGK
                :****:.****  ****  **** ****:****************.***: :    **.:

XM_011595       KNKGQLTKSPLAQMEEERREHVAKMKKMEMEMEQVFEMKVKEKVQKLKDSEAELQRRHEQ
NM_001788       KNKGQLTKSPLAQMEEERREHVAKMKKMEMEMEQVFEMKVKEKVQKLKDSEAELQRRHEQ
pnut            ARLS--NKNPLTQMEEEKREHEQKMKKMEAEMEQVFDMKVKEKMQKLRDSELELARRHEE
                 . .  .*.**:*****:***  ****** ******:******:***:*** ** ****:

XM_011595       MKKNLEAQHKELEEKRRQFEDEKANWEAQQRILEQQNSSRTLEKN-----------KKKG
NM_001788       MKKNLEAQHKELEEKRRQFEDEKANWEAQQRILEQQNSSRTLEKN-----------KKKG
pnut            RKKALELQIRELEEKRREFEREKKEWEDVNHVTLEELKRRSLGANSSTDNVDGKKEKKKK
                 ** ** * :*******:** ** :**  :::  :: . *:*  *           ***

XM_011595       KIF
NM_001788       KIF
pnut            GLF
                 :*
```

FIGURE 25. Expression of Peanut in mammalian tissues

FIGURE 25A. Real-time PCR analysis of Peanut expression in wildtype mouse tissues



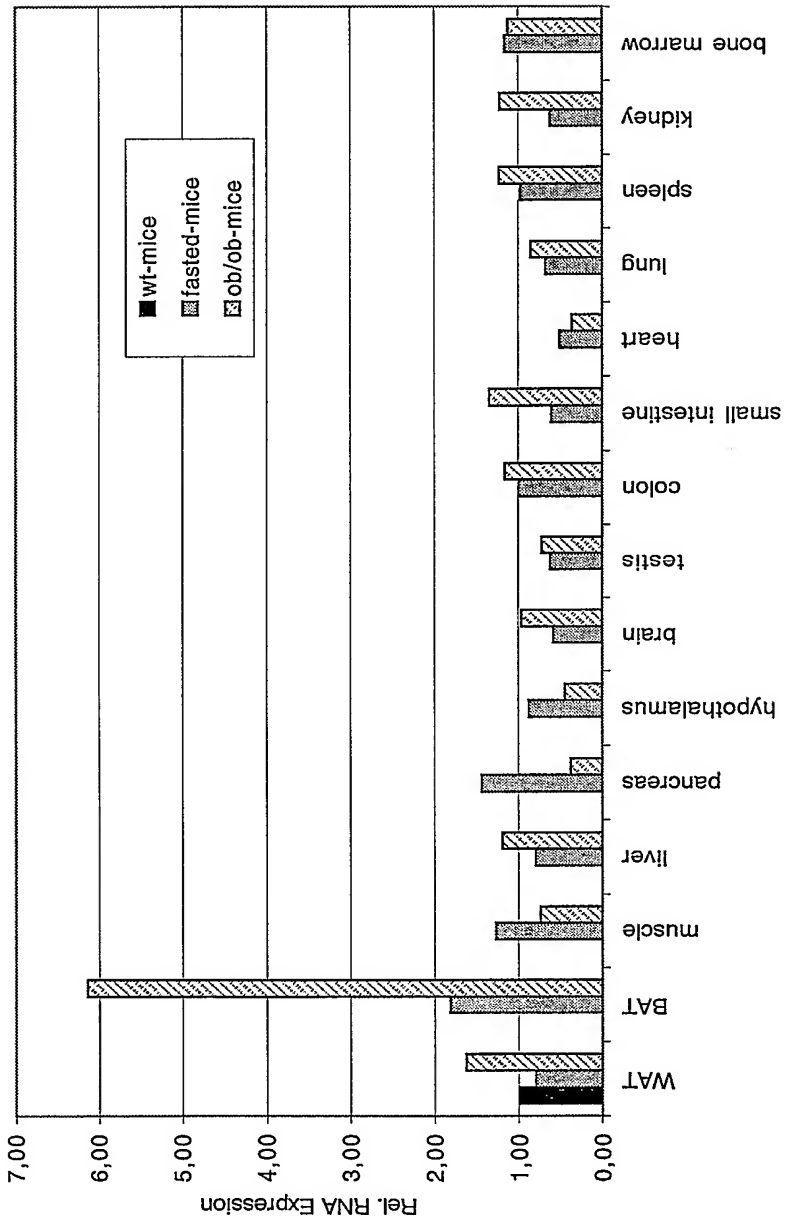


FIGURE 25B. Real-time PCR analysis of Peanut expression in different mouse models

FIGURE 26. Triglyceride content of a pyruvate kinase (Gadfly Accession Number CG7069) mutant

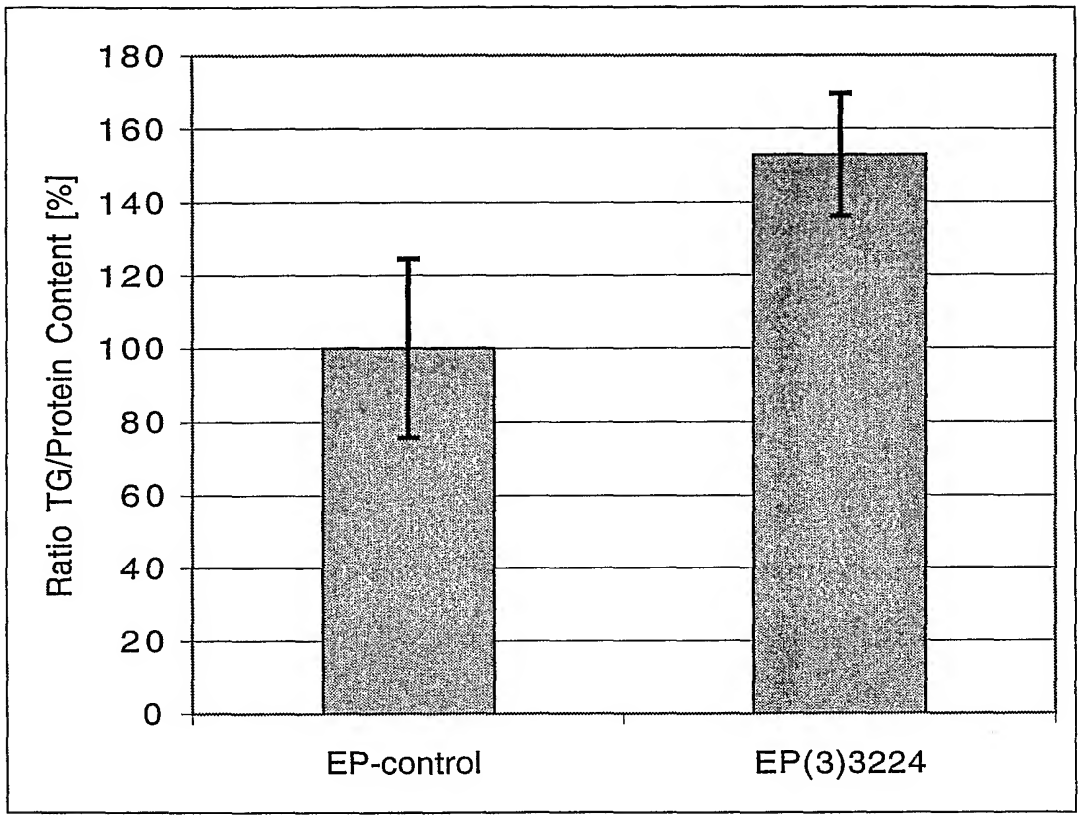


FIGURE 27. Molecular organisation of the pyruvate kinase gene (GadFly Accession Number CG7069)

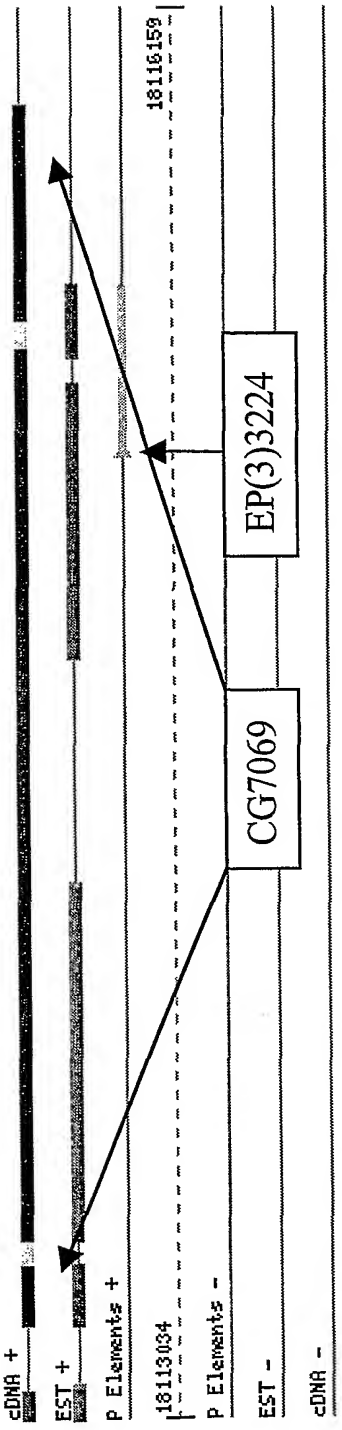


FIGURE 28: HUMAN HOMOLOG OF CG7069

FIGURE 28A. BLASTP result for CG7069 (Gadfly Accesssion Number)

Homology to human gene ref XM_037768; ref XP_037768.1 protein

```
>ref|XP_037768.1| (XM_037768) pyruvate kinase, muscle [Homo sapiens]
gb|AAH00481.1|AAH00481 (BC000481) pyruvate kinase, muscle [Homo sapiens]
gb|AAH07640.1|AAH07640 (BC007640) pyruvate kinase, muscle [Homo sapiens]
gb|AAH07952.1|AAH07952 (BC007952) pyruvate kinase, muscle [Homo sapiens]
Length = 531

Score =  410 bits (1043), Expect = e-113
Identities = 209/412 (50%), Positives = 284/412 (68%), Gaps = 2/412 (0%)


Query: 1   MRVVRMNFSHGSHEYHCQTIQAARKAIAMYVEQTGLPRTLAIALDTKGPEIRTGKLAGGN 60
           M V R+NFSHG+HEYH +TI+  R A   +     L R +A+ALDTKGPEIRTG L  G+
Sbjct: 69  MNVARLNFSHGTHEYHAETIKNVRTATESFASDPILYRPVAVALDTKGPEIRTG-LIKGS 127

Query: 61  DRAEIELKTGDKVTLSTKKEMADKSNKDNIYVDYQRLPQLVKPGNRVFVDDGLIALIVKE 120
             AE+ELK G   + ++       +K +++ +++DY+ + ++V+ G++++VDDGLI+L VK+
Sbjct: 128 GTAEVELKKGATLKITLDNAYMEKCDENILWLDYKNICKVVEVGSKIYVDDGLISLQVKQ 187

Query: 121 SKGDEVICQVENGGKLGSHKGINLPGVPVDLPSVTEKDKQDLKFGAEQKVDMIFASFIRD 180
              D ++ +VENGG LGS KG+NLPG  VDLP+V+EKD QDLKFG EQ VDM+FASFIR
Sbjct: 188 KGADFLVTEVENGGSLGSKKGVNLPGAAVDLPAVSEKDIQDLKFGVEQDVDMVFASFIRK 247

Query: 181 ANALKEIRQVLGPAGACIKIISKIENHQGLVNIDDIIRESDGIMVARGDMGIEIPTEDVP 240
           A+ + E+R+VLG  G  IKIISKIENH+G+   D+I+  SDGIMVARGD+GIEIP E V
Sbjct: 248 ASDVHEVRKVLGEKGKNIKIISKIENHEGVRRFDEILEASDGIMVARGDLGIEIPAEKVF 307

Query: 241 LAQKSIVAKCNKVGKPVICATQMMESMTNKPRPTRAEASDVANAIFDGCDAVMLSGETAK 300
           LAQK ++ +CN+ GKPVICATQM+ESM  KPRPTRAE SDVANA+ DG D +MLSGETAK
Sbjct: 308 LAQKMMIGRCNRAGKPVICATQMLESMIKKPRPTRAEGSDVANAVLDGADCIMLSGETAK 367

Query: 301 GKYPVECVQCMARICAKVEAVLWYESLQNSLKREIRTSAADHISAVTTAIAEAATVGQAR 360
           G YP+E V+    I  + EA +++  L   L+R +    +D   A      EA+    +
Sbjct: 368 GDYPLEAVRMQHLIAREAEAAIYHLQLFEELRR-LAPITSDPTEATAVGAVEASFKCCSG 426

Query: 361 AIVVASPCSMVAQMVSHMRPPCPIVMLTGNESEAAQSLLFRGIYPLLVEEMV 412
           AI+V +     A  V+  RP  PI+ +T N   A Q+ L+RGI+P+L ++ V
Sbjct: 427 AIIVLTKSGRSAHQVARYRPRAPIIAVTRNPQTARQAHLYRGIFPVLCKDPV 478
```

FIGURE 28B: Predicted nucleotide sequence encoding human pyruvate kinase, muscle (SEQ ID NO:27)

```
   1 cggcggcccg cagcgggata accttgaggc tgaggcagtg gctccttgca cagcagctgc
  61 acgcgccgtg gctccggatc tcttcgtctt tgcagcgtag cccgagtcgg tcagcagccg
 121 gaggacctca gcagccatgt cgaagcccca tagtgaagcc gggactgcct tcattcagac
 181 ccagcagctg cacgcagcca tggctgacac attcctggag cacatgtgcc gcctggacat
 241 tgattcacca cccatcacag cccggaacac tggcatcatc tgtaccattg gcccagcttc
 301 ccgatcagtg gagacgttga aggagatgat taagtctgga atgaatgtgg ctcgtctgaa
```

```
 361 cttctctcat ggaactcatg agtaccatgc ggagaccatc aagaatgtgc gcacagccac
 421 ggaaagcttt gcttctgacc ccatcctcta ccggcccgtt gctgtggctc tagacactaa
 481 aggacctgag atccgaactg ggctcatcaa gggcagcggc actgcagagg tggagctgaa
 541 gaagggagcc actctcaaaa tcacgctgga taacgcctac atggaaaagt gtgacgagaa
 601 catcctgtgg ctggactaca agaacatctg caaggtggtg gaagtgggca gcaagatcta
 661 cgtggatgat gggcttattt ctctccaggt gaagcagaaa ggtgccgact tcctggtgac
 721 ggaggtggaa aatggtggct ccttgggcag caagaagggt gtgaaccttc ctggggctgc
 781 tgtggacttg cctgctgtgt cggagaagga catccaggat ctgaagtttg gggtcgagca
 841 ggatgttgat atggtgtttg cgtcattcat ccgcaaggca tctgatgtcc atgaagttag
 901 gaaggtcctg ggagagaagg gaaagaacat caagattatc agcaaaatcg agaatcatga
 961 gggggttcgg aggtttgatg aaatcctgga ggccagtgat gggatcatgg tggctcgtgg
1021 tgatctaggc attgagattc ctgcagagaa ggtcttcctt gctcagaaga tgatgattgg
1081 acggtgcaac cgagctggga agcctgtcat ctgtgctact cagatgctgg agagcatgat
1141 caagaagccc ccgcccactc gggctgaagg cagtgatgtg gccaatgcag tcctggatgg
1201 agccgactgc atcatgctgt ctggagaaac agccaaaggg gactatcctc tggaggctgt
1261 gcgcatgcag cacctgattg cccgtgaggc agaggctgcc atctaccact tgcaattatt
1321 tgaggaactc cgccgcctgg cgcccattac cagcgacccc acagaagcca ccgccgtggg
1381 tgccgtggag gcctccttca agtgctgcag tggggccata atcgtcctca ccaagtctgg
1441 caggtctgct caccaggtgg ccagataccg cccacgtgcc cccatcattg ctgtgacccg
1501 gaatccccag acagctcgtc aggcccacct gtaccgtggc atcttccctg tgctgtgcaa
1561 ggacccagtc caggaggcct gggctgagga cgtggacctc cgggtgaact ttgccatgaa
1621 tgttgggtac gtggctggag caggggctag agcctagagg agcttgggga tgcttgagca
1681 ttggccacca acctcccttc tcttcctcca ggcaaggccc gaggcttctt caagaaggga
1741 gatgtggtca ttgtgctgac cggatggcgc cctggctccg gcttcaccaa caccatgcgt
1801 gttgttcctg tgccgtgatg gaccccagag cccctcctcc agcccctgtc ccacccccct
1861 cccccagccc atccattagg ccagcaacgc ttgtagaact cactctgggc tgtaacgtgg
1921 cactggtagg ttgggacacc agggaagaag atcaacgcct cactgaaaca tggctgtgtt
1981 tgcagcctgc tctagtggga cagcccagag cctggctgcc ccatcatgtg gccccaccca
2041 atcaagggaa gaaggaggaa tgctggactg gaggcccctg gagccagatg gcaagagggt
2101 gacagcttcc tttcctgtgt gtactctgtc cagttccttt agaaaaaatg gatgcccaga
2161 ggactcccaa ccctggcttg gggtcaagaa acagccagca agagttaggg gtccttaggg
2221 cactgggctg ttgttccatt gaagccgact ctggccctgg cccttacttg cttctctagc
2281 tctctaggcc tctccagttt gcacctgtcc ccaccctcca ctcagctgtc ctgcagcaaa
2341 cactccaccc tccaccttcc atttccccca ctactgcagc acctccaggc ctgttgctat
2401 agagcctacc tgtatgtaat aaa
```

FIGURE 28C: Predicted amino acid sequence of human human pyruvate kinase, muscle, M1 isozyme (SEQ ID NO:28)

```
  1 mskphseagt afiqtqqlha amadtflehm crldidsppi tarntgiict igpasrsvet
 61 lkemiksgmn varlnfshgt heyhaetikn vrtatesfas dpilyrpvav aldtkgpeir
121 tglikgsgta evelkkgatl kitldnayme kcdenilwld yknickvvev gskiyvddgl
181 islqvkqkga dflvteveng gslgskkgvn lpgaavdlpa vsekdiqdlk fgveqdvdmv
241 fasfirkasd vhevrkvlge kgknikiisk ienhegvrrf deileasdgi mvargdlgie
301 ipaekvflaq kmmigrcnra gkpvicatqm lesmikkprp traegsdvan avldgadcim
361 lsgetakgdy pleavrmqhl iareaeaaiy hlqlfeelrr lapitsdpte atavgaveas
421 fkccsgaiiv ltksgrsahq varyrprapi iavtrnpqta rqahlyrgif pvlckdpvqe
481 awaedvdlrv nfamnvgkar gffkkgdvvi vltgwrpgsg ftntmrvvpv p
```

FIGURE 28D: Predicted amino acid sequence of human human pyruvate kinase, muscle, M2 isozyme (SEQ ID NO:29)

```
  1 mskphseagt afiqtqqlha amadtflehm crldidsppi tarntgiict igpasrsvet
```

```
 61 lkemiksgmn varlnfshgt heyhaetikn vrtatesfas dpilyrpvav aldtkgpeir
121 tglikgsgta evelkkgatl kitldnayme kcdenilwld yknickvvev gskiyvddgl
181 islqvkqkga dflvteveng gslgskkgvn lpgaavdlpa vsekdiqdlk fgveqdvdmv
241 fasfirkasd vhevrkvlge kgknikiisk ienhegvrrf deileasdgi mvargdlgie
301 ipaekvflaq kmmigrcnra gkpvicatqm lesmikkprp traegsdvan avldgadcim
361 lsgetakgdy pleavrmqhl iareaeaaiy hlqlfeelrr lapitsdpte atavgaveas
421 fkccsgaiiv ltksgrsahq varyrprapi iavtrnpqta rqahlyrgif pvlckdpvqe
481 awaedvdlrv nfamnvgkar gffkkgdvvi vltgwrpgsg ftntmrvvpv p
```

FIGURE 28E: Predicted nucleotide sequence encoding human pyruvate kinase, liver and RBC (PKLR) (SEQ ID NO:30)

```
   1 gcagccccag gcccacactg aaagcatgtc gatccaggag aacatatcat ccctgcagct
  61 tcggtcatgg gtctctaagt cccaaagaga cttagcaaag tccatcctga ttggggctcc
 121 aggagggcca gcggggtatc tgcggcgggc cagtgtggcc caactgaccc aggagctggg
 181 cactgccttc ttccagcagc agcagctgcc agctgctatg gcagacacct tcctggaaca
 241 cctctgccta ctggacattg actccgagcc cgtggctgct cgcagtacca gcatcattgc
 301 caccatcggg ccagcatctc gctccgtgga gcgcctcaag gagatgatca aggccgggat
 361 gaacattgcg cgactcaact tctcccacgg ctcccacgag taccatgctg agtccatcgc
 421 caacgtccgg gaggcggtgg agagctttgc aggttcccca ctcagctacc ggcccgtggc
 481 catcgccctg gacaccaagg gaccggagat ccgcactggg atcctgcagg ggggtccaga
 541 gtcggaagtg gagctggtga agggctccca ggtgctggtg actgtggacc ccgcgttccg
 601 gacgcggggg aacgcgaaca ccgtgtgggt ggactacccc aatattgtcc gggtcgtgcc
 661 ggtgggggggc cgcatctaca ttgacgacgg gctcatctcc ctagtggtcc agaaaatcgg
 721 cccagaggga ctggtgaccc aagtggagaa cggcggcgtc ctgggcagcc ggaagggcgt
 781 gaacttgcca ggggcccagg tggacttgcc cgggctgtcc gagcaggacg tccgagacct
 841 gcgcttcggg gtggagcatg gggtggacat cgtctttgcc tcctttgtgc ggaaagccag
 901 cgacgtggct gccgtcaggg ctgctctggg tccggaagga cacggcatca agatcatcag
 961 caaaattgag aaccacgaag gcgtgaagag gtttgatgaa atcctggagg tgagcgacgg
1021 catcatggtg gcacgggggg acctaggcat cgagatccca gcagagaagg ttttcctggc
1081 tcagaagatg atgattgggc gctgcaactt ggcgggcaag cctgttgtct gtgccacaca
1141 gatgctggag agcatgatta ccaagccccg gccaacgagg gcagagacaa gcgatgtcgc
1201 caatgctgtg ctggatgggg ctgactgcat catgctgtca ggggagactg ccaagggcaa
1261 cttccctgtg gaagcggtga agatgcagca tgcgattgcc cgggaggcag aggccgcagt
1321 gtaccaccgg cagctgtttg aggagctacg tcgggcagcg ccactaagcc gtgatcccac
1381 tgaggtcacc gccattggtg ctgtggaggc tgccttcaag tgctgtgctg ctgccatcat
1441 tgtgctgacc acaactggcc gctcagccca gcttctgtct cggtaccgac ctcgggcagc
1501 agtcattgct gtcacccgct ctgcccaggc tgcccgccag gtccacttat gccgaggagt
1561 cttccccttg ctttaccgtg aacctccaga agccatctgg gcagatgatg tagatcgccg
1621 ggtgcaattt ggcattgaaa gtggaaagct ccgtggcttc ctccgtgttg gagacctggt
1681 gattgtggtg acaggctggc gacctggctc cggctacacc aacatcatga gggtgctaag
1741 catatcctga gacgcccctc ccccctctgg cccagcctac ccttgtaccc catcccttcc
1801 tccccagtct acgttctcca gcccacaccc ctccaaagcc ccacctttaa gtcctctctt
1861 ctctattcct gaccctccct acctgaggcc tatctgagac tataactgtc atctagcccc
1921 ttcgaggttg ccccttcccc atctccattt cacacaggtc ctgaaagtct gtgtccaatt
1981 atgcactggc cacccaacag caccaattgt acattctctg catccaatct gctcagcagg
2041 ccctaagatg ccttgagtct ttaatcccaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
2101 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
2161 aaaaaaaaa
```

FIGURE 28F: Predicted amino acid sequence of human human PKLR (SEQ ID NO:31)

```
1 msiqenissl qlrswvsksq rdlaksilig apggpagylr rasvaqltqe lgtaffqqqq
```

```
 61 lpaamadtfl ehlclldids epvaarstsi iatigpasrs verlkemika gmniarlnfs
121 hgsheyhaes ianvreaves fagsplsyrp vaialdtkgp eirtgilqgg pesevelvkg
181 sqvlvtvdpa frtrgnantv wvdypnivrv vpvggriyid dglislvvqk igpeglvtqv
241 enggvlgsrk gvnlpgaqvd lpglseqdvr dlrfgvehgv divfasfvrk asdvaavraa
301 lgpeghgiki iskienhegv krfdeilevs dgimvargdl gieipaekvf laqkmmigrc
361 nlagkpvvca tqmlesmitk prptraetsd vanavldgad cimlsgetak gnfpveavkm
421 qhaiareaea avyhrqlfee lrraaplsrd ptevtaigav eaafkccaaa iivltttgrs
481 aqllsryrpr aaviavtrsa qaarqvhlcr gvfpllyrep peaiwaddvd rrvqfgiesg
541 klrgflrvgd lvivvtgwrp gsgytnimrv lsis
```

FIGURE 29. CLUSTAL W (1.7) Protein Sequence Alignment Analysis

```
pk3_h2          MSKPHSEAGTAFIQTQQLHAAMADTFLEHMCRLDIDSPPITARNTGIICTIGPASRSVET
pk3_h           MSKPHSEAGTAFIQTQQLHAAMADTFLEHMCRLDIDSPPITARNTGIICTIGPASRSVET
pk3_m           MPKPHSEAGTAFIQTQQLHAAMADTFLEHMCRLDIDSAPITARNTGIICTIGPASRSVEM
pk3_dro         ------------------------------------------------------------


pk3_h2          LKEMIKSGMNVARLNFSHGTHEYHAETIKNVRTATESFASDPILYRPVAVALDTKGPEIR
pk3_h           LKEMIKSGMNVARLNFSHGTHEYHAETIKNVRTATESFASDPILYRPVAVALDTKGPEIR
pk3_m           LKEMIKSGMNVARLNFSHGTHEYHAETIKNVREATESFASDPILYRPVAVALDTKGPEIR
pk3_dro         --------MRVVRMNFSHGSHEYHCQTIQAARKAIAMYVEQTGLPRTLAIALDTKGPEIR
                        *.*.*:*****:****.:**: .* *   :..:. * *.:*:**********

pk3_h2          TGLIKGSG-TAEVELKKGATLKITLDNAYMEKCDENILWLDYKNICKVVEVGSKIYVDDG
pk3_h           TGLIKGSG-TAEVELKKGATLKITLDNAYMEKCDENILWLDYKNICKVVEVGSKIYVDDG
pk3_m           TGLIKGSG-TAEVELKKGATLKITLDNAYMEKCDENILWLDYKNICKVVEVGSKIYVDDG
pk3_dro         TGKLAGGNDRAEIELKTGDKVTLSTKKEMADKSNKDNIYVDYQRLPQLVKPGNRVFVDDG
                ** : *..  **:***.* .:.:: .:   :*.::: :::**:.: ::*: *.:::****

pk3_h2          LISLQVKQKGADFLVTEVENGGSLGSKKGVNLPGAAVDLPAVSEKDIQDLKFGVEQDVDM
pk3_h           LISLQVKQKGADFLVTEVENGGSLGSKKGVNLPGAAVDLPAVSEKDIQDLKFGVEQDVDM
pk3_m           LISLQVKEKGADFLVTEVENGGSLGSKKGVNLPGAAVDLPAVSEKDIQDLKFGVEQDVDM
pk3_dro         LIALIVKESKGDEVICQVENGGKLGSHKGINLPGVPVDLPSVTEKDKQDLKFGAEQKVDM
                **:* **:. .* :: :*****.***:**:****..****:*:*** ******.**.***

pk3_h2          VFASFIRKASDVHEVRKVLGEKGKNIKIISKIENHEGVRRFDEILEASDGIMVARGDLGI
pk3_h           VFASFIRKASDVHEVRKVLGEKGKNIKIISKIENHEGVRRFDEILEASDGIMVARGDLGI
pk3_m           VFASFIRKAADVHEVRKVLGEKGKNIKIISKIENHEGVRRFDEILEASDGIMVARGDLGI
pk3_dro         IFASFIRDANALKEIRQVLGPAGACIKIISKIENHQGLVNIDDIIRESDGIMVARGDMGI
                :******.*  ::*:*:***  *  **********:*: .:*:*:. **********:**

pk3_h2          EIPAEKVFLAQKMMIGRCNRAGKPVICATQMLESMIKKPRPTRAEGSDVANAVLDGADCI
pk3_h           EIPAEKVFLAQKMMIGRCNRAGKPVICATQMLESMIKKPRPTRAEGSDVANAVLDGADCI
pk3_m           EIPAEKVFLAQKMMIGRCNRAGKPVICATQMLESMIKKPRPTRAEGSDVANAVLDGADCI
pk3_dro         EIPTEDVPLAQKSIVAKCNKVGKPVICATQMMESMTNKPRPTRAEASDVANAIFDGCDAV
                ***:*.* **** ::.:**:.**********:*** :********.******::**.*.:

pk3_h2          MLSGETAKGDYPLEAVRMQNLIAREAEAAIYHLQLFEELRR-LAPITSDPTEATAVGAVE
pk3_h           MLSGETAKGDYPLEAVRMQHLIAREAEAAIYHLQLFEELRR-LAPITSDPTEATAVGAVE
pk3_m           MLSGETAKGDYPLEAVRMQHLIAREAEAAIYHLQLFEELRR-LAPITSDPTEAAAVGAVE
pk3_dro         MLSGETAKGKYPVECVQCMARICAKVEAVLWYESLQNSLKREIRTSAADHISAVTTAIAE
                *********.**:*.*:    *. :.**.::: .* :.*:* : . ::*  .*.:.. .*

pk3_h2          ASFKCCSGAIIVLTKSGRSAHQVARYRPRAPIIAVTRNPQTARQAHLYRGIFPVLCKDPV
pk3_h           ASFKCCSGAIIVLTKSGRSAHQVARYRPRAPIIAVTRNPQTARQAHLYRGIFPVLCKDPV
pk3_m           ASFKCCSGAIIVLTKSGRSAHQVARYRPRAPIIAVTRNPQTARQAHLYRGIFPVLCKDAV
pk3_dro         AATVGQARAIVVASPCSMVAQMVSHMRPPCPIVMLTGNESEAAQSLLFRGIYPLLVEEMV
                *:    : **:* : ..  *: *:: ** .**: :* * . * *: *:***:*:* :: *

pk3_h2          QEAWAEDVDLRVNFAMNVGKARGFFK--KGDVVIVLTGWRPGSGF-TNTMRVVPVP----
pk3_h           QEAWAEDVDLRVNFAMNVGKARGFFK--KGDVVIVLTGWRPGSGF-TNTMRVVPVP----
pk3_m           LNAWAEDVDLRVNLAMDVGKARGFFK--KGDVVIVLTGWRPGSGF-TNTMRVVPVP----
pk3_dro         IGSFNFRRIMQSGLKL-MGKMDILEPGQKGSVVLVNAMSAEKITFRLFTIRQQTKEERDQ
                  ::     :: .: : :**   :    **.**:* :       *   *:*  .
```

```
pk3_h2      ------------------------------------------------------------
pk3_h       ------------------------------------------------------------
pk3_m       ------------------------------------------------------------
pk3_dro     DERCRKLALEQSCKERAEKEECRKLQQAEECQKQKLAKKCKQFEEKQKVCPKKNDTPKND

pk3_h2      ------------------------------------------------------------
pk3_h       ------------------------------------------------------------
pk3_m       ------------------------------------------------------------
pk3_dro     CPKKDCPKKECPKQDDEISKCRQMQEAEAEERKCKEEFEQMCKLAEEKRKEAEKCRKADE

pk3_h2      ------------------------------------------------------------
pk3_h       ------------------------------------------------------------
pk3_m       ------------------------------------------------------------
pk3_dro     ERRKEEAEKCRKLEEDRKCKLAEEKKRNEEELKIIEAEVAKLEAAEKAKRLKEEEKKKEE

pk3_h2      ------------------------------------------------------------
pk3_h       ------------------------------------------------------------
pk3_m       ------------------------------------------------------------
pk3_dro     LMKCKQRNEAKKKREEAERCKRKERERELAEMENKWKQVAEKRKRKKAAEMCRKIEDAKE

pk3_h2      ---------------------------------
pk3_h       ---------------------------------
pk3_m       ---------------------------------
pk3_dro     KAAAESADKILKAVCEKLKQSLSDPDKSKKGKK
```

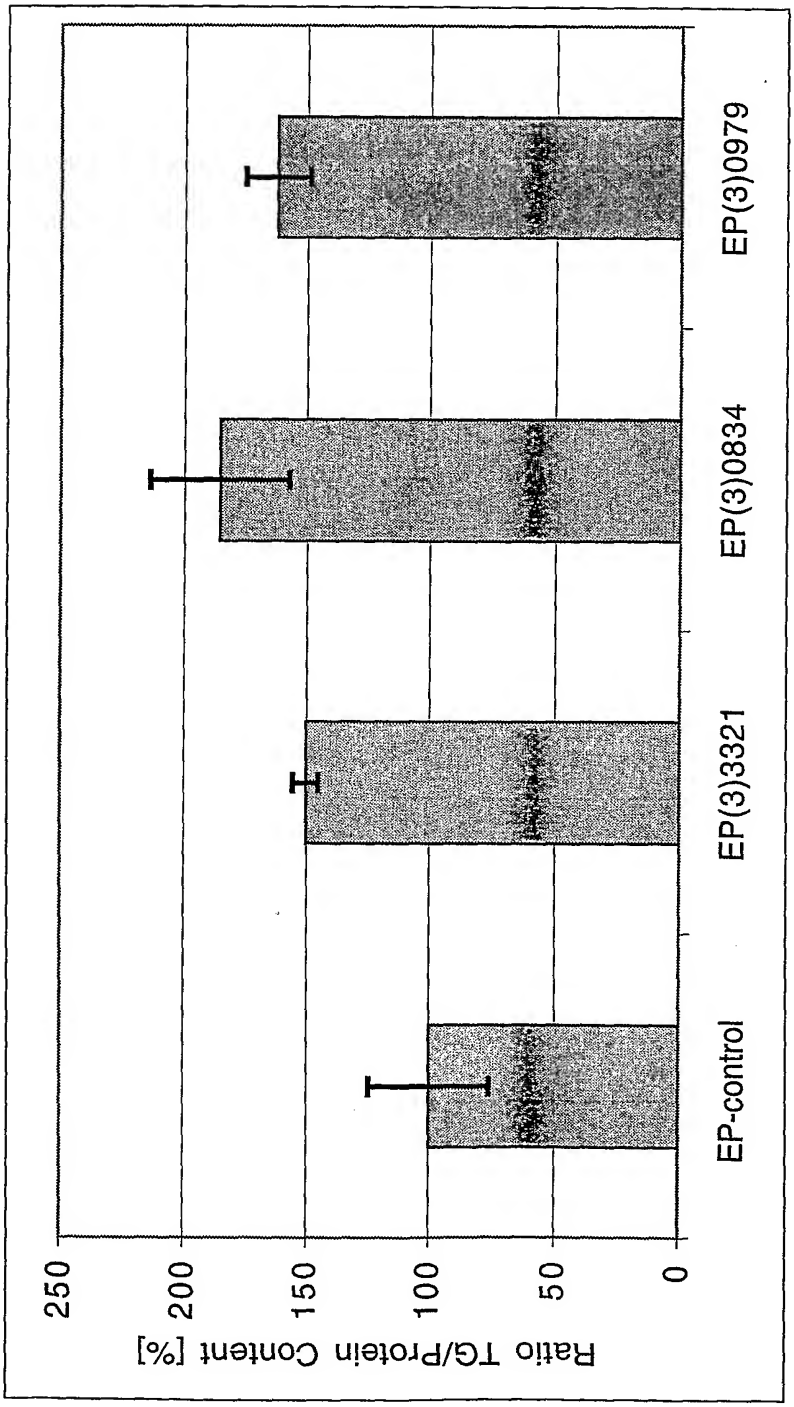


FIGURE 30. Triglyceride content of *Calreticulin* (crc; Gadfly Acession Number CG9429) mutants

FIGURE 31. Molecular organisation of *Calrreticulin* (crc; Gadfly Accession Number CG9429)

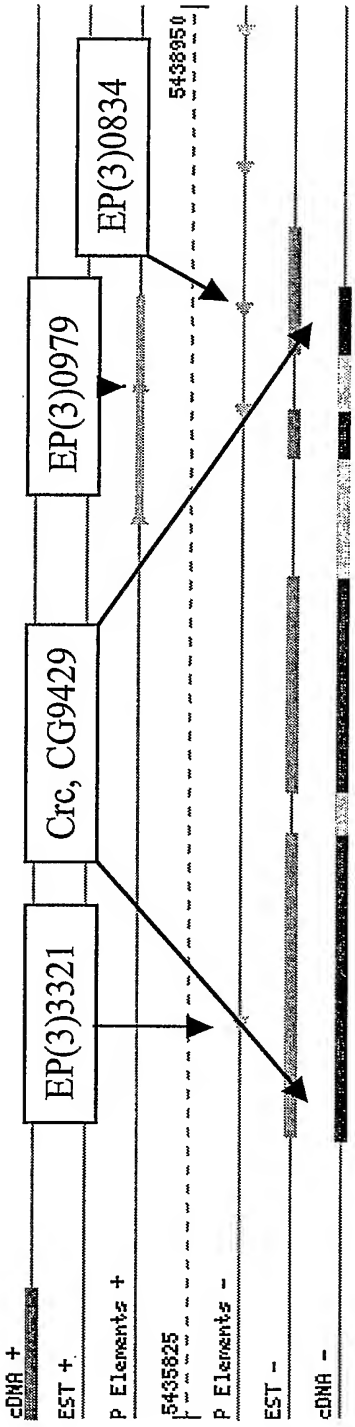


FIGURE 32: HUMAN HOMOLOG OF CG9429 (Calreticulin, crc)

FIGURE 32A. BLASTP search result for crc (Gadfly Accession Number CG9429)

```
ref|NP_004334.1| (NM_004343) calreticulin precursor; Sicca syndrome antigen A
 (autoantigen Ro; calreticulin); autoantigen Ro [Homo sapiens]
Length = 417

Score =  575 bits (1467), Expect = e-163
Identities = 269/404 (66%), Positives = 317/404 (77%), Gaps = 5/404 (1%)

Query: 6   TVIVLLATVGFISAE--VYLKENF-DNENWEDTWIYSKHPGKEFGKFVLTPGTFYNDAEA 62
           +V +LL  +G   AE  VY KE F D + W   WI SKH   +FGKFVL+ G FY D E
Sbjct: 4   SVPLLLGLLGLAVAEPAVYFKEQFLDGDGWTSRWIESKHKS-DFGKFVLSSGKFYGDEEK 62

Query: 63  DKGIQTSQDARFYAASRKFDGFSNEDKPLVVQFSVKHEQNIDCGGGYVKLFDCSLDQTDM 122
           DKG+QTSQDARFYA S  F+ FSN+ + LVVQF+VKHEQNIDCGGGYVKLF  SLDQTDM
Sbjct: 63  DKGLQTSQDARFYALSASFEPFSNKGQTLVVQFTVKHEQNIDCGGGYVKLFPNSLDQTDM 122

Query: 123 HGESPYEIMFGPDICGPGTKKVHVIFSYKGKNHLISKDIRCKDDVYTHFYTLIVRPDNTY 182
           HG+S Y IMFGPDICGPGTKKVHVIF+YKGKN LI+KDIRCKDD +TH YTLIVRPDNTY
Sbjct: 123 HGDSEYNIMFGPDICGPGTKKVHVIFNYKGKNVLINKDIRCKDDEFTHLYTLIVRPDNTY 182

Query: 183 EVLIDNEKVESGNLEDDWDFLAPKKIKDPTATKPEDWDDRATIPDPDDKKPEDWDKPEHI 242
           EV IDN +VESG+LEDDWDFL PKKIKDP A+KPEDWD+RA I DP D KPEDWDKPEHI
Sbjct: 183 EVKIDNSQVESGSLEDDWDFLPPKKIKDPDASKPEDWDERAKIDDPTDSKPEDWDKPEHI 242

Query: 243 PDPDATKPEDWDDEMDGEWEPPMIDNPEFKGEWQPKQLDNPNYKGAWEHPEIANPEYVPD 302
           PDPDA KPEDWD+EMDGEWEPP+I NPE+KGEW+P+Q+DNP+YKG W HPEI NPEY PD
Sbjct: 243 PDPDAKKPEDWDEEMDGEWEPPVIQNPEYKGEWKPRQIDNPDYKGTWIHPEIDNPEYSPD 302

Query: 303 DKLYLRKEICTLGFDLWQVKSGTIFDNVLITDDVELAAKAAAEVKN-TQAGEKKMKEAQD 361
             +Y       LG DLWQVKSGTIFDN LIT+D   A +   E    T+A EK+MK+ QD
Sbjct: 303 PSIYAYDNFGVLGLDLWQVKSGTIFDNFLITNDEAYAEEFGNETWGVTKAAEKQMKDKQD 362

Query: 362 EVQRKKDEEEAKKASDKDDEDEDDDDEEKDDESKQDKDQSEHDE 405
           E QR K+EEE KK  ++++ ++ +DDE+KD++ + ++D+ E +E
Sbjct: 363 EEQRLKEEEEDKKRKEEEEAEDKEDDEDKDEDEEDEEDKEEDEE 406
```

FIGURE 32B: Predicted nucleotide sequence encoding human Calreticulin (SEQ ID NO:32)

```
   1 gtccgtactg cagagccgct gccggagggt cgttttaaag ggccgcgttg ccgcccctc
  61 ggcccgccat gctgctatcc gtgccgctgc tgctcggcct cctcggcctg gccgtcgccg
 121 agcccgccgt ctacttcaag gagcagtttc tggacggaga cgggtggact tcccgctgga
 181 tcgaatccaa acacaagtca gattttggca aattcgttct cagttccggc aagttctacg
 241 gtgacgagga gaaagataaa ggtttgcaga caagccagga tgcacgcttt tatgctctgt
 301 cggccagttt cgagcctttc agcaacaaag gccagacgct ggtggtgcag ttcacggtga
 361 aacatgagca gaacatcgac tgtgggggcg gctatgtgaa gctgtttcct aatagtttgg
 421 accagacaga catgcacgga gactcagaat acaacatcat gtttggtccc gacatctgtg
 481 gccctggcac caagaaggtt catgtcatct tcaactacaa gggcaagaac gtgctgatca
 541 acaaggacat ccgttgcaag gatgatgagt ttacacacct gtacacactg attgtgcggc
 601 cagacaacac ctatgaggtg aagattgaca acagccaggt ggagtccggc tccttggaag
 661 acgattggga cttcctgcca cccaagaaga taaaggatcc tgatgcttca aaaccggaag
 721 actgggatga gcgggccaag atcgatgatc ccacagactc caagcctgag gactgggaca
 781 agcccgagca tatccctgac cctgatgcta agaagcccga ggactgggat gaagagatgg
 841 acggagagtg ggaacccccа gtgattcaga accctgagta caagggtgag tggaagcccc
 901 ggcagatcga caacccagat tacaagggca cttggatcca cccagaaatt gacaaccccg
 961 agtattctcc cgatcccagt atctatgcct atgataactt tggcgtgctg ggcctggacc
1021 tctggcaggt caagtctggc accatctttg acaacttcct catcaccaac gatgaggcat
1081 acgctgagga gtttggcaac gagacgtggg gcgtaacaaa ggcagcagag aaacaaatga
1141 aggacaaaca ggacgaggag cagaggctta aggaggagga agaagacaag aaacgcaaag
1201 aggaggagga ggcagaggac aaggaggatg atgaggacaa agatgaggat gaggaggatg
1261 aggaggacaa ggaggaagat gaggaggaag atgtccccgg ccaggccaag gacgagctgt
1321 agagaggcct gcctccaggg ctggactgag gcctgagcgc tcctgccgca gagcttgccg
1381 cgccaaataa tgtctctgtg agactcgaga actttcattt ttttccaggc tggttcggat
1441 ttggggtgga ttttggtttt gttcccctcc tccactctcc cccacccccт ccccgccctt
1501 ttttttttтt ttttaaact ggtattttat cctttgattc tccttcagcc ctcacccctg
1561 gttctcatct ttcttgatca acatcttttc ttgcctctgt gcccctтctc tcatctctta
1621 gctcccctcc aacctggggg gcagtggtgt ggagaagcca caggcctgag atttcatctg
1681 ctctccttcc tggagcccag aggagggcag cagaaggggg tggtgtctcc aaccccccag
1741 cactgaggaa gaacggggct cttctcattt caccctccc tttctcccct gcccccagga
1801 ctgggccact tctgggtggg gcagtgggtc ccagattggc tcacactgag aatgtaagaa
1861 ctacaaacaa aatttctatt aaattaaatt ttgtgtctc
```

FIGURE 32C: Predicted amino acid sequence of human Calreticulin (SEQ ID NO:33)

```
  1 mllsvpllllg llglavaepa vyfkeqfldg dgwtsrwies khksdfgkfv lssgkfygde
 61 ekdkglqtsq darfyalsas fepfsnkgqt lvvqftvkhe qnidcgggyv klfpnsldqt
121 dmhgdseyni mfgpdicgpg tkkvhvifny kgknvlinkd irckddefth lytlivrpdn
181 tyevkidnsq vesgsleddw dflppkkikd pdaskpedwd erakiddptd skpedwdkpe
241 hipdpdakkp edwdeemdge weppviqnpe ykgewkprqi dnpdykgtwi hpeidnpeys
301 pdpsiyaydn fgvlgldlwq vksgtifdnf litndeayae efgnetwgvt kaaekqmkdk
361 qdeeqrlkee eedkkrkeee eaedkedded kdedeedeed keedeeedvp gqakdel
```

FIGURE 32D: Predicted nucleotide sequence encoding human Calreticulin 2 (SEQ ID NO:34)

```
   1 agcggagagg cgcagagaga gctgggagct aaggggtggc ggcgaccgga agcgcagtgc
  61 acacccccat ggcccgggct ttggtccagt tctgggccat atgcatgctg cgagtggcgc
 121 tggctaccgt ctatttccaa gaggaatttc tagacggaga gcattggaga aaccgatggt
 181 tgcagtccac caatgactcc cgatttgggc attttagact ttcgtcgggc aagttttatg
 241 gtcataaaga gaaagataaa ggtctgcaaa ccactcagaa tggccgattc tatgccatct
 301 ctgcacgctt caaaccgttc agcaataaag ggaaaactct ggttattcag tacacagtaa
 361 aacatgagca gaagatggac tgtggagggg gctacattaa ggtctttcct gcagacattg
 421 accagaagaa cctgaatgga aaatcgcaat actatattat gtttggaccc gatatttgtg
 481 gatttgatat caagaaagtt catgttattt tacatttcaa gaataagtat cacgaaaaca
 541 agaaactgat caggtgtaag gttgatggct tcacacacct gtacactcta attttaagac
 601 cagatctttc ttatgatgtg aaaattgatg gtcagtcaat tgaatccggc agcatagagt
 661 acgactggaa cttaacatca ctcaagaagg aaacgtcccc ggcagaatcg aaggattggg
 721 aacagactaa agacaacaaa gcccaggact gggagaagca ttttctggac gccagcacca
 781 gcaagcagag cgactggaac ggtgacctgg atggggactg gccagcgccg atgctccaga
 841 agcccccgta ccaggatggc ctgaaaccag aaggtattca taaagacgtc tggctccacc
 901 gtaagatgaa gaataccgac tatttgacgc agtatgacct ctcagaattt gagaacattg
 961 gtgccattgg cctggagctt tggcaggtga gatctggaac cattttgat aactttctga
1021 tcacagatga tgaagagtat gcagataatt ttggcaaggc cacctggggc gaaaccaagg
1081 gtccagaaag ggagatggat gccatacagg ccaaggagga aatgaagaag gcccgcgagg
1141 aagaggagga agagctgctg tcgggaaaaa ttaacaggca cgaacattac ttcaatcaat
1201 ttcacagaag gaatgaactt tagtgatccc cattggatat aaggatgact ggtaaaatct
1261 cattgctact ttaatctaaa aaaaaaaaaa aaa
```

FIGURE 32E: Predicted amino acid sequence of human Calreticulin 2 (SEQ ID NO:35)

```
   1 maralvqfwa icmlrvalat vyfqeefldg ehwrnrwlqs tndsrfghfr lssgkfyghk
  61 ekdkglqttq ngrfyaisar fkpfsnkgkt lviqytvkhe qkmdcgggyi kvfpadidqk
 121 nlngksqyyi mfgpdicgfd ikkvhvilhf knkyhenkkl irckvdgfth lytlilrpdl
 181 sydvkidgqs iesgsieydw nltslkkets paeskdweqt kdnkaqdwek hfldastskq
 241 sdwngdldgd wpapmlqkpp yqdglkpegi hkdvwlhrkm kntdyltqyd lsefenigai
 301 glelwqvrsg tifdnflitd deeyadnfgk atwgetkgpe remdaiqake emkkareeee
 361 eellsgkinr hehyfnqfhr rnel
```

FIGURE 33. CLUSTAL W (1.82) Protein Sequence Alignment Analysis

```
crc Dm        MMWCKTVIVLLATVGFISAEVYLKENFDN-ENWEDTWIYSKHPGKEFGKFVLTPGTFYND
crc Hs        MLLSVPLLLGLLGLAVAEPAVYFKEQFLDGDGWTSRWIESKHKS-DFGKFVLSSGKFYGD
MGC26577 Hs   MARALVQFWAICMLRVALATVYFQEEFLDGEHWRNRWLQSTNDS-RFGHFRLSSGKFYGH

crc Dm        AEADKGIQTSQDARFYAASRKFDGFSNEDKPLVVQFSVKHEQNIDCGGGYVKLFDCSLDQ
crc Hs        EEKDKGLQTSQDARFYALSASFEPFSNKGQTLVVQFTVKHEQNIDCGGGYVKLFPNSLDQ
MGC26577 Hs   KEKDKGLQTTQNGRFYAISARFKPFSNKGKTLVIQYTVKHEQKMDCGGGYIKVFPADIDQ

crc Dm        TDMHGESPYEIMFGPDICGPGTKKVHVIFSYKGKNHLISKDIRCKDDVYTHFYTLIVRPD
crc Hs        TDMHGDSEYNIMFGPDICGPGTKKVHVIFNYKGKNVLINKDIRCKDDEFTHLYTLIVRPD
MGC26577 Hs   KNLNGKSQYYIMFGPDICGFDIKKVHVILHFKNKYHENKKLIRCKVDGFTHLYTLILRPD

crc Dm        NTYEVLIDNEKVESGNLEDDWDFLAPKKIKDPTATKPEDWDDRATIPDPDDKKPEDWDKP
crc Hs        NTYEVKIDNSQVESGSLEDDWDFLPPKKIKDPDASKPEDWDERAKIDDPTDSKPEDWDKP
MGC26577 Hs   LSYDVKIDGQSIESGSIEYDWNLTSLKKETSPAESK--DWEQTK------DNKAQDWEK-

crc Dm        EHIPDPDATKPEDWDDEMDGEWEPPMIDNPEFKGEWQPKQLDNPNYKGAWEHPEIANPEY
crc Hs        EHIPDPDAKKPEDWDEEMDGEWEPPVIQNPEYKGEWKPRQIDNPDYKGTWIHPEIDNPEY
MGC26577 Hs   -HFLDASTSKQSDWNGDLDGDWPAPMLQKPPYQDGLKPEGIH----KDVWLHRKMKNTDY

crc Dm        VPDDKLYLRKEICTLGFDLWQVKSGTIFDNVLITDDVELAAKAAAEVK-NTQAGEKKMKE
crc Hs        SPDPSIYAYDNFGVLGLDLWQVKSGTIFDNFLITNDEAYAEEFGNETWGVTKAAEKQMKD
MGC26577 Hs   LTQYDLSEFENIGAIGLELWQVRSGTIFDNFLITDDEEYADNFGKATWGETKGPEREMDA

crc Dm        AQDEVQRKKDEEEAKKASDKDDED--EDDDDEEKDDESKQDKDQSE--------HDEL
crc Hs        KQDEEQRLKEEEEDKKRKEEEEAEDKEDDEDKDEDEEDEEDKEEDEEEDVPGQAKDEL
MGC26577 Hs   IQAK-------EEMKKAREEEEEELLSGKINRHEHYFNQFHR------------RNEL
```